Are People Naïve Probability Theorists? A Further Examination of the Probability Theory + Variation Model

CHRISTOPHER R. FISHER* and CHRISTOPHER R. WOLFE
Miami University, Oxford, OH, USA

ABSTRACT
Two experiments tested predictions derived from the Probability Theory + Variation (PTV) model. PTV model assumes that judgments follow probability theory, but systematic errors arise from noise in the judgments. Experiment 1 compared the PTV model to a configurational weighted averaging model in joint probability judgment and found more support for the PTV model in diagnostic cases. Specifically, noise was negatively correlated with semantic coherence and conjunction and disjunction fallacies increased when order effects produced more noise in conjunctions and disjunctions. Consistent with both models, judgments adhered stochastically to the addition law. Contrary to the integration rules of the PTV model, we failed to find increased noise in disjunctions compared to conjunctions. Experiment 2 tested predictions of the PTV model for conditional probability judgment. Consistent with the PTV model, noise was negatively correlated with semantic coherence in conditional probabilities and judgments adhered stochastically to Bayes’ theorem. Conversion errors were generally more prevalent than conditional reversals, a finding that is not fully consistent with the PTV model. In general, the quantitative fit of the PTV model was relatively better for overlapping and subset problems compared to identical, independent and mutually exclusive problems, especially for semantic coherence. Copyright © 2014 John Wiley & Sons, Ltd.

KEY WORDS probability judgment; mathematical modeling; stochastic models; conjunction fallacy

INTRODUCTION
Converging evidence over the past 40 years suggests that human judgment deviates systematically from the norms of probability theory. In their seminal paper, Tversky and Kahneman (1983) demonstrated that people reliably commit the conjunction fallacy—erroneously judging the conjunction of two events as more probable than one of its components, a finding that has been frequently replicated (e.g., Costello, 2009a; Wolfe, Fisher, Reyna, & Hu, 2012; Yates & Carlson, 1986). In the original study, they described a fictitious character named Linda, who was described in terms of a feminist stereotype. On the basis of the description, participants judged the probability that Linda is a bank teller and the probability that Linda is a feminist bank teller. Approximately 85% of participants judged feminist bank teller as more probable than bank teller, even though feminist bank teller is a subset of bank tellers. The fact that such a simple rule was so easily violated cast considerable doubt that human judgment is consistent with probability theory. Casting further doubt, other non-normative phenomena have been observed, such as underconfidence and overconfidence (Erev, Wallsten, & Budescu, 1994), subadditivity (Tversky & Koehler, 1994), the disjunction fallacy (Fisk, 2002), base rate neglect in Bayesian inference (Kahneman & Tversky, 1973), and the conversion error (Wolfe, 1995) just to name a few. Collectively, this corpus of research suggests that people do not judge probabilities in a manner consistent with probability theory, leading many to conclude that there are systematic irrationalities in human cognition (Stanovich & West, 2000).

In response, various non-normative psychological models have been proposed to account for these findings. One of the earliest explanations was that humans use simple rules or heuristics that work well in certain situations but can lead to systematic biases in others (Tversky & Kahneman, 1983). According to the representativeness heuristic, the conjunction fallacy occurs in the Linda problem because feminist bank teller is more representative or similar to Linda than bank teller. Because representativeness is not constrained by the rules of probability theory, errors such as the conjunction fallacy can occur. However, subsequent research suggests that representativeness is not the underlying cause of the conjunction fallacy. For example, the conjunction fallacy rate is similar for standard problems (e.g., the Linda problem) and probability combination problems in which the representativeness heuristic does not apply (Gavanski & Roskos-Ewoldsen, 1991). The probability combination problems used unfamiliar content (e.g., aliens from another planet), and the component probabilities were yoked to judgments from the standard problem. The similar rates of the conjunction fallacy suggest that probabilities are integrated in a non-normative fashion irrespective of representativeness.

Other non-normative integration models have been proposed to account for the conjunction fallacy (Abelson, Leddo, & Gross, 1987; Fantino, Kulik, Stolarz-Fantino, & Wright, 1997; Yates & Carlson, 1986; Wolfe, 1995). Perhaps the simplest among these models is the averaging model. According to the averaging model, conjunctions are formed by averaging the component probabilities (Fantino et al., 1997). One shortcoming of the averaging model is that it almost always predicts the occurrence of a conjunction fallacy because an average falls between the component probabilities. The configurational weighted average (CWA) model circumvents this problem by incorporating noise in the component probabilities, allowing it to produce normative responses as well (Nilsson, Winman, Juslin, & Hansson, 2009).
As an alternative approach, some stochastic models begin with the assumption that judgments are more or less normative, but stipulate that noise or random variability in the cognitive system produces systematic errors. One of the earliest examples was a stochastic model of calibration used to account for a perplexing phenomenon in which underconfidence and overconfidence was observed in similar probability judgment tasks (Erev et al., 1994). According to the model, probability judgments are well calibrated, but noise produces underconfidence or overconfidence, depending on the circumstances. The model proposes a two-stage judgment process in which a covert judgment is made followed by an overt judgment. The covert judgment is based on a subjective level of confidence in the truth of a statement, which is represented on an unbounded scale and is a function of the true probability and an additive error component. The overt judgment is made by mapping the covert judgment onto the interval [0, 1] to produce a probability judgment. This process can produce underconfidence or overconfidence, depending on the manner in which objective probabilities are defined. The defining feature of this model is that both underconfidence and overconfidence can be explained without assuming that the underlying judgments are poorly calibrated.

Building upon these ideas, the Probability Theory + Variation (PTV) model was recently proposed to account for other judgment phenomena, such as the conjunction fallacy (Costello, 2009a), the disjunction fallacy (Costello, 2009b), and subadditivity (Costello & Watts, 2012). What distinguishes the PTV model from the stochastic calibration model is that it models how probabilities are integrated to form judgments of compound events, such as conjunctions and disjunctions. The PTV model assumes that probability judgments are integrated coherently according to the rules of probability theory, but the underlying cognitive system is noisy. Although noise in the PTV model represents random variability, the model predicts systematic errors, such as the conjunction fallacy. It is important to note that the PTV model does not postulate that people explicitly follow probability theory. If they did, errors should not be observed. Instead, the model postulates that cognitive system operates in a manner consistent with probability theory at a computational level, except the processes are perturbed with noise.

Although many decision theorists may consider the PTV model to be implausible in light of the multitude of reported departures from probability judgment, the PTV model has been successful in accounting for a wide range of judgment phenomena. For example, the PTV model can account for two robust determinants of the conjunction fallacy and disjunction fallacies—the component probabilities and the conditional support between them (Costello, 2009a; Costello, 2009b). The rate of the conjunction fallacy is highest when one component probability is low and the other is high, followed by two high probability components (Fisk, 2002; Gavanski & Roskos-Ewoldsen, 1991; Nilsson et al., 2009). The lowest rate occurs when both component probabilities are low. Another robust determinant is the conditional support one event provides to the other (Crisp & Feeney, 2009; Tversky & Kahneman, 1983). Positive conditional support occurs when \( P(\text{B|A}) > P(\text{B}) \), whereas negative conditional support occurs when \( P(\text{B|A}) < P(\text{B}) \). Because the model predicts that the disjunction and conjunction fallacies are both determined by the component probabilities and conditional support, it correctly predicts a strong correlation between conjunction and disjunction fallacy rates (Costello, 2009b). In addition, the PTV model can account for double-conjunction and double-disjunction fallacies (Costello, 2009a; Costello, 2009b) and has a high degree of quantitative fit to fallacy rates (Costello, 2009a).

Our goal in the present research was to subject the PTV model to rigorous empirical testing. Experiment 1 compared the PTV model to the CWA model (Nilsson et al., 2009) in the domain of joint probability judgment. The CWA model was selected for comparison because it makes divergent predictions regarding the relationship between noise and certain errors, such as the conjunction fallacy. Experiment 1 examined whether judgments adhere stochastically to the addition law, as predicted by both models (Costello & Watts, 2012). In addition, Experiment 1 also tested a critical property of the PTV model in which the noise in disjunctions is greater than the noise in conjunctions. Experiment 2 tested the predictions of the PTV model in the domain of conditional probability judgment. In particular, we tested whether increased noise is associated with increased errors, as predicted by the PTV model. We also tested whether judgments adhere stochastically to Bayes’ theorem. Before detailing the model predictions, we provide a formal description of the PTV model and CWA model in the next section.

**MODELS**

**The PTV model**

Recall that the PTV model assumes that judgments follow the rules of probability theory but are perturbed with noise (Costello, 2009a). Let \( k \in \{A, B, A \cap B, A \cup B, A|B, B|A\} \) and let \( S(k) \in [0,1] \) be a random variable representing the reported subjective probability of event \( k \). The subjective probability can be decomposed into a true probability component and an additive error component:

\[
S(k) = P(k) + e_k
\]

(1)

where \( P(k) \) is the true judgment that is consistent with probability theory and \( e_k \) is an error term that can assume positive or negative values. The expectation of the subjective probability equals the values given by probability theory:

\[
E[S(k)] = P(k)
\]

(2)

Predictions from the model are derived through the manipulation of the rules of probability theory and adding error terms each of the components. To illustrate, first consider the conjunction rule:

\[
P(A \cap B) = P(A)P(B|A)
\]

(3)

Next, error terms are added to define the conjunction rule in terms of the model:
where the error terms are assumed to be independent. Equation (4) indicates that the subjective conjunctive probability is a function of the component and conditional probabilities and their respective error terms. Assuming \( P(A) > P(B), \) a conjunction fallacy occurs when the following inequality is true:

\[
S(B) < S(A \cap B)
\]  

(5)

Expressing Equation (5) in terms of the following model:

\[
P(B) + e_B < [P(A) + e_A][P(B|A) + e_{B|A}]
\]

(6)

In Equation (6), the predicted conjunction fallacy rate is equal to the probability that the aforementioned inequality is true. An examination of Equation (6) reveals an important implication of the model: the rate of the conjunction fallacy increases when the error terms increase. In the absence of noise, the conjunction fallacy should never occur.

The CWA model

For joint probability judgment, we compared the PTV model with the CWA model (Nilsson et al., 2009), which makes divergent predictions in certain cases. Unlike the PTV model, the CWA model is inherently non-normative in the sense that it produces conjunction and disjunction fallacies in the absence of noise. According to the CWA model, joint probability judgments are formed through a CWA of noisy component probabilities. Assuming \( S(A) > S(B), \)

\[
S(A \cap B) = (1 - w)S(A) + wS(B)
\]

(7)

\[
S(A \cup B) = wS(A) + (1 - w)S(B)
\]

(8)

where \( w = .80. \) Nilsson et al. (2009) argued that additive integration of the component probabilities can produce accurate joint probabilities in the presence of noise in the environment and cognitive system when \( w \) is set to .80 a priori.

PREDICTIONS FOR EXPERIMENT 1

Relationship between fallacies

The model predictions are summarized in Table 1. As previously noted, the conjunction and disjunction fallacies increase as one component becomes less likely and the other becomes more likely (e.g., Fisk, 2002; Nilsson et al., 2009). One prediction that follows from this finding is that the rate of conjunction and disjunction fallacies should be correlated. Both models make this prediction, which has been supported previously (e.g., Costello, 2009b).

Noise and errors

In some cases, the PTV model and CWA model make divergent predictions regarding the relationship between noise and errors (Table 1). The expected rank order of judgments for each model is provided in the following to explicate the derivations of the predictions. Assuming \( S(A) > S(B), \) the PTV model predicts the following rank order:

\[
S(A \cup B) \geq S(A) > S(B) \geq S(A \cap B)
\]

(9)

By contrast, the CWA model predicts a different rank order:

\[
S(A) > S(A \cup B) > S(A \cap B) > S(B)
\]

(10)

According to both models, deviations from the predicted rank orders are due to noise. The PTV model makes a consistent prediction regarding the relationship between noise and errors: increasing noise will increase errors. By contrast, the CWA model makes different predictions depending on the specific error under investigation.

Whereas the PTV model predicts that conjunction fallacies will increase with noise, the CWA model predicts that conjunction fallacies will decrease with noise. Inspection of the expected rank order of the CWA model reveals that in the absence of noise, a conjunction fallacy is predicted: \( S(A \cap B) > S(B). \) Thus, increasing noise will produce more normative responses by chance. In support of the CWA model, Nilsson et al. (2009) found that averaging multiple judgments from the same person on the same problem increased rather than decreased the conjunction and disjunction fallacy. However, others have argued that this averaging procedure does not reduce noise, but instead reduces the proportional difference (Costello & Watts, 2012). In light of this potential problem, we used a straightforward, alternative analysis to evaluate whether increased noise is associated with increased errors. This alternative analysis evaluates the

Table 1. Summary of model predictions

<table>
<thead>
<tr>
<th>Prediction</th>
<th>PTV</th>
<th>CWA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation conjunction and disjunction fallacy</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Correlation between noise and errors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conjunction fallacy</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Disjunction fallacy</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Double-conjunction fallacy</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Double-disjunction fallacy</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Minimum conjunction error</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Maximum disjunction error</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Semantic coherence</td>
<td>&lt;</td>
<td>&gt;</td>
</tr>
</tbody>
</table>

Note: Divergent predictions are italicized. PTV, Probability Theory + Variation; CWA, configural weighted average; NP, no prediction.
relationship between a person’s error rates and the amount of noise in his or her judgments.\(^1\)

The same predictions can be derived for the disjunction fallacy, which occurs when the disjunction is judged as less probable than the larger component probability, \(S(A)\):

\[
S(A) > S(A \cup B) \quad (11)
\]

As before, the PTV model predicts that the disjunction fallacy will increase with noise, whereas the CWA model predicts that the disjunction fallacy will decrease with noise.

Next, we turn to the double-conjunction and double-disjunction fallacies. A double-conjunction fallacy occurs when the conjunction is judged as more probable than the high probability component, \(P(A)\):

\[
S(A) < S(A \cap B) \quad (12)
\]

A double-disjunction fallacy occurs when the disjunction is judged as less probable than the low-probability component, \(P(B)\):

\[
S(B) > S(A \cup B) \quad (13)
\]

As before, the PTV model predicts that the double-conjunction and double-disjunction fallacies will increase with noise. The CWA model also predicts that double-conjunction and double-disjunction fallacies should increase with noise. To see why this is the case, consider the double-conjunction fallacy as an example. The CWA model implies \(S(A) > S(A \cap B)\) in the absence of noise. However, in the presence of noise, the expected rank order may reverse for any given judgment to produce a double-conjunction fallacy.

In addition to conjunction and disjunction fallacies, we investigated alternative coherence benchmarks. Two of these alternative coherence benchmarks are based on the Fréchet inequalities (Fréchet, 1951), which we term as the minimum conjunction error and the maximum disjunction error. Unlike the conjunction fallacy, a minimum conjunction error occurs when a conjunction is too low (Wolfe & Reyna, 2010). As an example, suppose \(S(A) = S(B) = .60\). According to probability theory, the sum of all disjoint events must equal 1. Because \(S(A) + S(B) = 1.20\), probability theory requires that the minimum conjunction must be .20. In general, a minimum conjunction error occurs when

\[
\max(0, S(A) + S(B) - 1) > S(A \cap B) \quad (14)
\]

As the minimum conjunction error represents a departure from probability theory, the PTV model predicts that it should increase with noise. Similarly, the CWA model predicts that the minimum conjunction error will increase with noise because in the absence of noise the model predicts conjunctions that are too large rather than too small.

Unlike the disjunction fallacy, a maximum disjunction error occurs when a disjunction is too high rather than too low. According to the addition law in Equation (15), the disjunction cannot be larger than the sum of the components: \(S(A) + S(B)\).

\[
S(A \cup B) = S(A) + S(B) - S(A \cap B) \quad (15)
\]

In general, the maximum disjunction error occurs when

\[
\min(1, S(A) + S(B)) < S(A \cup B) \quad (16)
\]

As the maximum disjunction error represents a departure from probability theory, the PTV model predicts that it should increase with noise. Similarly, the CWA model predicts that the maximum disjunction error will increase with noise because in the absence of noise the model predicts disjunctions that are too small rather too large.

Previous studies have shown that people are sensitive to the semantic content in probability judgment problems (Wolfe & Reyna, 2010; Wolfe, Fisher, & Reyna, 2013). For this reason, we use semantic coherence to test whether the model can capture peoples’ sensitivity to semantic content. Semantic coherence is a more stringent coherence benchmark that assesses whether a set of judgments maps onto the qualitative set relationship implied by a given problem (Wolfe & Reyna, 2010). There are five qualitative relationships between two events, \(A\) and \(B\): identical (e.g., \(H2O\) and \(water\)), mutually exclusive (e.g., \(bee\) and \(wasp\)), subset (e.g., \(cat\) and \(mammal\)), independent (e.g., heads on a coin \(flip\) and \(rain\)), and overlapping (e.g., \(feminist\) and \(bank\) \(teller\)). Consider the simple case of two identical events, \(H2O\) and \(water\). To be semantically coherent, a set of judgments must satisfy the following constraints: \(P(A) = P(B) = P(A \cap B) = P(A \cup B) > 0\). Because the rules for semantic coherence are derived from probability theory, the PTV model predicts that increased noise should be associated with lower semantic coherence (for details, refer to Wolfe & Reyna, 2010). By contrast, the CWA model predicts the opposite.

### The addition law

The PTV and CWA model both predict that judgments should adhere stochastically to the addition law of probability theory. To see why this is the case for the CWA model, substitute Equations (7) and (8) into Equation (17) and simplify. Although noise will cause individual sets of judgments will deviate from 0, they should be distributed accordingly:

\[
S(A) + S(B) - S(A \cap B) - S(A \cup B) \sim F(0, \sigma^2) \quad (17)
\]

where \(F\) has a mean of zero. Support for this prediction was found in Costello and Watts (2012). For both models, a corollary of this prediction is that the noise in individual sets of judgments should be correlated with the absolute deviation from zero in Equation (17).
Noise in joint probabilities

One implication of the PTV model described in Costello (2009a, 2009b) is that noise in disjunctions should be greater than noise in conjunctions. According to the PTV model, conjunctions and disjunctions are a function of noisy component and conditional probabilities, which are integrated according to the rules of probability theory. Because the component and conditional probabilities are independent random variables, their variances are defined as follows:

\[
\text{Var}[S(A \cap B)] = \text{Var}[S(A)]S(B|A)
\]

(18)

\[
\text{Var}[S(A \cup B)] = \text{Var}[S(A)] + \text{Var}[S(B)] + \text{Var}[S(A)]S(B|A]
\]

(19)

In comparing Equation (18) with Equation (19), it becomes clear that the variance in the conjunction is a subset of the variance in the disjunction. It stands to reason that more noise should be observed in disjunctions, unless \( \text{Var}[S(A)] + \text{Var}[S(B)] = 0 \). The CWA model does not make a simple prediction about the relative noise in conjunctions and disjunctions.

Order effects

The conjunction fallacy has been found to be higher when the conjunction is rated first followed by its components (Stolarz-Fantino, Fantino, Zizzo, & Wen, 2003). According to the PTV model, attention modulates the amount of noise in the component and conditional probabilities (Costello, 2009a). When the component probabilities are judged first, they are maintained in attention when the conjunction is subsequently judged. As a result, the components become relatively fixed, which, in turn, decreases the chance of a conjunction fallacy. In other words, attention decreases the error terms in Equation (6), thereby making the conjunction fallacy less likely to occur. By contrast, less attention is given to the component and conditional probabilities when the conjunction is judged first. In this case, the judgments are more prone to random variation, which increases the chance of a conjunction fallacy. Because the conjunctions and disjunctions are a function of noisy inputs, this prediction extends to the other errors and semantic coherence. Although Nilsson et al. (2009) did not provide an account of order effects, the attentional mechanism proposed by Costello (2009a) appears to be consistent with the CWA model. For this reason, we extend the attentional mechanism to the CWA model to permit the comparison of the models. To the extent that an order effect is observed, the predictions for the PTV model and CWA model mirror those for the relationship between noise and errors (Table 1). The models make divergent predictions for conjunction fallacies, disjunction fallacies, and semantic coherence for the reasons detailed in Errors and Noise.

Copyright © 2014 John Wiley & Sons, Ltd.

EXPERIMENT 1

Participants

Participants were 61 introductory psychology students at a Midwestern University, who participated for partial course credit. Consistent with previous studies at this University, participants were disproportionately white and female.

Materials

A pilot study was conducted to develop the problems used in Experiment 1 and Experiment 2. In the interest of brevity, the description of the pilot study is merged with Experiment 1 because both used the same procedures. The problems used in the pilot study were designed to systematically vary low and high component probability combinations, conditional support, and set types. Most problems were adopted from published studies and modified as necessary (e.g., Crisp & Feeney, 2009; Wolfe & Reyna, 2010; Wolfe et al., 2013; Wolfe & Fisher, 2013), whereas the remaining problems were developed specifically for this study. Each problem featured a short scenario followed by questions for \( P(A), P(B), P(A \cap B), \) and \( P(A \cup B) \). As an example, consider the following problem: “Steve is 50 years old and has a sedentary lifestyle. He is a movie buff. When he comes home from his job as a computer programmer, he likes to watch movies from his movie collection and eat his favorite ice cream: double fudge, chocolate chip with sprinkles.” The conjunction and disjunction were formed from the two component events: \( (A) \) Steve is obese and \( (B) \) Steve can do 50 push-ups. This problem was designed to have one high \((A)\) and one low-probability event \((B)\), negative conditional support, and depict overlapping sets. The final 34 problems were selected from a larger set of problems developed in the pilot study based on the variety in the component probabilities, conditional support, set type, and having sufficient variability between judgments at time 1 and time 2. In a few cases, minor adjustments were made before using the piloted materials in the main experiments. In addition, we included a total of 22 filler probability judgment problems. Finally, an argumentation filler task consisted of a subset of 25 simple arguments adopted from Wolfe and Brit (2008).

Procedures

Participants completed the study individually on computers in groups ranging from one to five. A typical completion time ranged from 45 to 55 minutes. Participants completed two blocks of judgments consisting of 34 target problems and 11 filler problems presented in randomized order. The 34 target problems were presented in both blocks of judgments, but a different set of 11 filler items was used in each block. Each problem consisted of a short description followed by questions for \( P(A), P(B), P(A \cap B), \) and \( P(A \cup B) \), which were presented individually. For each judgment, participants entered a number between 0 and 100. The judgments were blocked by component probabilities and joint probabilities, with items randomized within each block. To test the order effects, judgment blocks of component and
joint probabilities were counterbalanced across participants. After completing the first block of judgments, participants completed the filler argumentation task to interfere with memory. After reading each argument, participants rated the argument in terms of personal agreement and argument strength on a Likert scale. Lastly, participants completed the second block of judgments.

RESULTS

Two participants were excluded because of a computer error. An additional five participants were excluded because they failed to complete experiment within the allotted time, resulting in a total of 54 participants. Before turning to our primary analyses, we compared the difference in mean judgments for time 1 and time 2 across problems and judgments. The difference in mean judgments (mean = −.005, SD = .039) was not statistically significant, *t*(135) = −1.63, *p* = .11. The small observed difference provides evidence that the judgments were noisy. According to the PTV and CWA models, the rate of conjunction and disjunction fallacies should be correlated because they are influenced by the same factors. Consistent with Costello (2009b), the correlation between the rate of conjunction and disjunction fallacies was *r*(32) = .86, *p* < .001 across problems.

Noise and errors

One critical property of the PTV model is that error rates should increase as noise in judgments increases. To test this property, we computed the correlations between each participant’s error and semantic coherence rates and a measure of overall judgment noise. Overall judgment noise was computed as the mean absolute difference between judgments at time 1 and time 2. For minimum conjunction errors and maximum disjunction errors, we restricted our analyses to the relevant cases in which the errors can occur. The restriction is S(A) + S(B) > 1 for the minimum conjunction error and S(A) + S(B) < 1 for the maximum disjunction error. In these cases, we weighted each participant’s scores according to the number of contributing observations (Pinto da Costa, 2011). The results are summarized in Table 2. Consistent with both models, double conjunction and double disjunction fallacies were both correlated with overall judgment noise. However, the negative correlation between semantic coherence and overall judgment noise uniquely supports the PTV model. The remaining correlations failed to reach statistical significance.

Order effects

The PTV model and our extension of the CWA model predict that judging joint probabilities before component probabilities will increase noise in the joint probabilities. To test this prediction, the mean absolute differences for the conjunctions and disjunctions were computed across all problems for each participant. Table 3 shows that judgment order increased noise for conjunctions in the predicted

<table>
<thead>
<tr>
<th>Error</th>
<th><em>r</em></th>
<th><em>p</em>-value</th>
<th>Mean rate</th>
<th>Standard deviation rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1 (N=54)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conjunction fallacy</td>
<td>.02</td>
<td>.89</td>
<td>.14</td>
<td>.08</td>
</tr>
<tr>
<td>Disjunction fallacy</td>
<td>.06</td>
<td>.68</td>
<td>.12</td>
<td>.07</td>
</tr>
<tr>
<td>Double-conjunction fallacy</td>
<td>.40</td>
<td>.003</td>
<td>.08</td>
<td>.06</td>
</tr>
<tr>
<td>Double-disjunction Fallacy</td>
<td>.48</td>
<td>&lt;.001</td>
<td>.08</td>
<td>.08</td>
</tr>
<tr>
<td>Minimum conjunction error†</td>
<td>−.06</td>
<td>.68</td>
<td>.17</td>
<td>.13</td>
</tr>
<tr>
<td>Maximum disjunction error</td>
<td>.20</td>
<td>.14</td>
<td>.12</td>
<td>.11</td>
</tr>
<tr>
<td>Semantic coherence</td>
<td>−.28</td>
<td>.04</td>
<td>.11</td>
<td>.07</td>
</tr>
<tr>
<td>Sum of errors</td>
<td>.34</td>
<td>.01</td>
<td>.54</td>
<td>.23</td>
</tr>
<tr>
<td>Experiment 2 (N=55)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum conditional error A</td>
<td>B</td>
<td>.18</td>
<td>.19</td>
<td>.26</td>
</tr>
<tr>
<td>Minimum conditional error B</td>
<td>A</td>
<td>.15</td>
<td>.27</td>
<td>.28</td>
</tr>
<tr>
<td>Conditional reversal</td>
<td>.31</td>
<td>.02</td>
<td>.08</td>
<td>.07</td>
</tr>
<tr>
<td>Conversion error</td>
<td>.17</td>
<td>.22</td>
<td>.14</td>
<td>.06</td>
</tr>
<tr>
<td>Semantic coherence</td>
<td>−.43</td>
<td>.001</td>
<td>.24</td>
<td>.09</td>
</tr>
<tr>
<td>Sum of errors</td>
<td>.11</td>
<td>.41</td>
<td>.41</td>
<td>.19</td>
</tr>
</tbody>
</table>

*Based on *N*=53.

<table>
<thead>
<tr>
<th>Error</th>
<th>Components first</th>
<th>Joint first</th>
<th><em>t</em>-test</th>
<th><em>p</em>-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conjunction fallacy</td>
<td>.12</td>
<td>.16</td>
<td>−1.97</td>
<td>.05</td>
</tr>
<tr>
<td>Disjunction fallacy</td>
<td>.10</td>
<td>.15</td>
<td>−2.68</td>
<td>.01</td>
</tr>
<tr>
<td>Double-conjunction fallacy</td>
<td>.08</td>
<td>.09</td>
<td>−0.84</td>
<td>.40</td>
</tr>
<tr>
<td>Double-disjunction fallacy</td>
<td>.06</td>
<td>.09</td>
<td>−1.50</td>
<td>.14</td>
</tr>
<tr>
<td>Minimum conjunction error*</td>
<td>.21</td>
<td>.14</td>
<td>1.84</td>
<td>.07</td>
</tr>
<tr>
<td>Maximum disjunction error</td>
<td>.11</td>
<td>.13</td>
<td>−0.89</td>
<td>.38</td>
</tr>
<tr>
<td>Semantic coherence</td>
<td>.10</td>
<td>.08</td>
<td>1.52</td>
<td>.14</td>
</tr>
<tr>
<td>Noise in conjunctions</td>
<td>.09</td>
<td>.13</td>
<td>−3.21</td>
<td>.002</td>
</tr>
<tr>
<td>Noise in disjunctions</td>
<td>.10</td>
<td>.13</td>
<td>−1.57</td>
<td>.12</td>
</tr>
<tr>
<td>Noise in conjunctions min*&lt;sup&gt;†&lt;/sup&gt;</td>
<td>.17</td>
<td>.25</td>
<td>−4.15</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Noise in disjunctions min*&lt;sup&gt;†&lt;/sup&gt;</td>
<td>.14</td>
<td>.21</td>
<td>−1.59</td>
<td>.12</td>
</tr>
<tr>
<td>Noise in conjunctions max&lt;sup&gt;†&lt;/sup&gt;</td>
<td>.12</td>
<td>.17</td>
<td>−1.67</td>
<td>.10</td>
</tr>
<tr>
<td>Noise in disjunctions max&lt;sup&gt;†&lt;/sup&gt;</td>
<td>.16</td>
<td>.20</td>
<td>−1.34</td>
<td>.19</td>
</tr>
</tbody>
</table>

* *N*=28 in components first and *N*=26 in joint first.
† Tests performed on restricted observations for cases in which minimum conjunction or maximum disjunction errors may occur.

direction. The increase for disjunctions was in the predicted direction, but not statistically significant. The most diagnostic result is the increase in conjunction and disjunction fallacies when the joint probabilities are judged first. This finding is consistent with the PTV model, but not with the CWA model. Except for the minimum conjunction error, the differences in error and semantic coherence rates were non-significant but in the direction predicted by the PTV model.

The addition law
Both models predict stochastic adherence to the addition law. To test whether the addition law holds, \(S(A) + S(B) - S(A \cap B)\) was computed for each participant’s judgments on each problem. Because of the high number of observations (54 participants × 34 problems × 3 judgments = 5136), we more precisely evaluated the predictions for the addition law with confidence intervals. There was a small but systematic bias from the predicted mean of 0, mean = .025, SD = .23, 95% CI [.015, .034]. Although the mean shows a systematic deviation from 0, it is a small deviation in comparison to the full range of possible values [-2, 2]. One corollary of the addition law is that the variability in the distribution should be associated with the amount of noise in the judgments. To test this prediction, a composite measure of noise was formed by summing the absolute deviations between time 1 and time 2 judgments for each participant on each problem. Because of non-normality in the distributions, Spearman’s rank correlation was used. As predicted by the model, the correlation between absolute difference and the composite noise was in the predicted direction, \(r = .29, 95\% \text{ CI}[.26, .32]\).

Noise in joint probabilities
The PTV model predicts that the noise should be greater in disjunctions than conjunctions because disjunctions are a function of more random variables. To test this prediction, the variances for the conjunctions and disjunctions were computed for each set of judgments (54 subjects × 34 problems = 1836). Contrary to the PTV model, the mean difference was essentially zero, mean difference = .001, 95% CI [-.005, .002]. The expected difference in conjunctive and disjunctive variances should approximate the sum of the unique variance terms in Equation (19): \(\text{Var}[S(A)] + \text{Var}[S(B)]\). The sum of the mean variances was .052, which is much larger than the observed difference.

Model fit
The PTV model was fit to aggregated data using the basic procedures described by Costello (2009a) (refer to Appendix 1 for details). In brief, the error and semantic coherence rates were estimated for each problem through simulation of the subjective probabilities. The mean and standard deviations of the aggregated judgments were allowed to vary within ±.02 of their observed values to adjust for sampling error while sufficiently constraining the model. The corresponding mean and standard deviations for conditional probability judgments were taken from Experiment 2. The results are summarized in Table 4. The mean absolute difference between predicted and observed rates was .07 (SD = .09), which was larger in comparison to the results reported in Costello (2009a). The correlation between predicted and observed rates was \(r = .47\). The model performed relatively well on independent and overlapping problems in comparison to identical, mutually exclusive, and subset problems. There was a general tendency to overestimate the rate of maximum disjunction errors and underestimate semantic coherence, which was particularly pronounced for identical problems.

DISCUSSION
In Experiment 1, we compared the PTV model to the CWA model. Several of the results were consistent with both models. For example, double-conjunction and double-disjunction fallacies increased with noise. In addition, judgments adhered to the addition law as predicted by both models. In cases where the models made divergent predictions, the PTV model generally received better support. For example, semantic coherence decreased as noise increased, which uniquely supports the PTV model. For the order effects, the conjunction and disjunction fallacies increased as predicted by the PTV model. Except for the minimum conjunction error, the other results were non-significant but in the direction predicted by the PTV model. This may be due to low statistical power to detect a medium effect size of \(d = .50\) (power = .44, two-tailed). Contrary to the PTV model, however, the variances (i.e., noise) for conjunctions and disjunctions were essentially identical. This result is at odds with integration rules of the PTV model.

Table 4. Quantitative predictions of the PTV model for joint probabilities in Experiment 1. Predicted rates are in parentheses

<table>
<thead>
<tr>
<th></th>
<th>CF</th>
<th>DF</th>
<th>DCF</th>
<th>DDF</th>
<th>MinC</th>
<th>MaxD</th>
<th>SC</th>
<th>Mean absolute difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identical</td>
<td>.01</td>
<td>.01</td>
<td>.09</td>
<td>.17</td>
<td>.08</td>
<td>.16</td>
<td>.14</td>
<td>.03</td>
</tr>
<tr>
<td>Independent</td>
<td>.27</td>
<td>.25</td>
<td>.20</td>
<td>.25</td>
<td>.05</td>
<td>.04</td>
<td>.04</td>
<td>.05</td>
</tr>
<tr>
<td>Mutually exclusive</td>
<td>.02</td>
<td>.11</td>
<td>.04</td>
<td>.11</td>
<td>.06</td>
<td>.06</td>
<td>.10</td>
<td>.06</td>
</tr>
<tr>
<td>Overlapping</td>
<td>.20</td>
<td>.19</td>
<td>.17</td>
<td>.19</td>
<td>.11</td>
<td>.08</td>
<td>.08</td>
<td>.04</td>
</tr>
<tr>
<td>Subset</td>
<td>.08</td>
<td>.12</td>
<td>.09</td>
<td>.12</td>
<td>.08</td>
<td>.12</td>
<td>.07</td>
<td>.04</td>
</tr>
<tr>
<td>Mean absolute difference</td>
<td>.05</td>
<td>.05</td>
<td>.06</td>
<td>.15</td>
<td>.06</td>
<td>.08</td>
<td>.07</td>
<td></td>
</tr>
</tbody>
</table>

CF, conjunction fallacy; DF, disjunction fallacy; DCF, double-conjunction fallacy; DDF, double-disjunction fallacy; MinC, minimum conjunction error; MaxD, maximum disjunction error; SC, semantic coherence.
model, which imply that disjunctions should have more noise than conjunctions.

The quantitative tests revealed mixed support for the PTV model. Although the PTV model performed relatively well on overlapping and independent problems, it had more difficulty accounting for identical, mutually exclusive and subset problems. In addition, the PTV model generally had trouble accounting for the maximum disjunction error and semantic coherence. The difficulty accounting for semantic coherence was particularly pronounced for problems featuring identical sets. One question that deserves attention is why the discrepancy was particularly pronounced for problems featuring identical

predictions for experiment 2

noise and errors

The goal of Experiment 2 was to extend tests of the PTV model to conditional probability judgment, using the same procedures as Experiment 1. For this reason, many of the predictions parallel those described for joint probability judgment. As with joint probability judgment, we test whether increased noise is associated with increased errors. One common error in conditional probability judgment is the conversion error (Reyna & Brainerd, 2008). The conversion error occurs when the conditional probabilities are erroneously judged to be equal—that is, when P(\(A|B\)) = P(\(B|A\)) and P(\(A\)) \neq P(\(B\)). Another error that can occur is the conditional reversal (Fisher & Wolfe, 2011). Formally, a conditional reversal occurs when P(\(A|B\)) < P(\(B|A\)) and P(\(A\)) > P(\(B\)).

In joint probability judgment, the minimum conjunction error occurs when a conjunction is too small. Because of the relationship between conditional and conjunctive probabilities, it is possible to assess a related error, which we call the minimum conditional error (Fisher & Wolfe, 2011). The definition of a conditional probability states that

\[
P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{20}
\]

Dividing both sides of 14 by P(\(A\)) and substituting equation (20), a minimum conditional error occurs when

\[
\max \left(0, \frac{S(A) + S(B) - 1}{S(A)} \right) > S(B|A) \tag{21}
\]

There are two minimum conditional errors because Equation (14) can be divided by P(\(B\)) as well.

Bayes’ theorem

The PTV model predicts that judgments should adhere stochastically to Bayes’ theorem, which is defined as

\[
P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{22}
\]

Multiplying both sides of Equation (22) by P(\(B\)) and expressing it in terms of subjective probabilities results in

\[
S(B)S(A|B) - S(A)S(B|A) \sim G(0, \sigma^2) \tag{23}
\]

Thus, the PTV model predicts a distribution with a mean of zero if judgments adhere stochastically to Baye’s theorem. In addition, the PTV model predicts that absolute deviations from zero should be correlated with the amount of judgment noise.

METHODS

Participants

Participants were 62 introductory psychology students at a Midwestern University, who completed the experiment for partial course credit.

Materials and procedures

Experiment 2 used the same materials and procedures as Experiment 1, with one exception. In Experiment 2, participants judged the probability of events A, B, A\(\cap\)B, and B\(\cap\)A instead of A, B, A\(\cap\)B, and A\(\cup\)B.

RESULTS

Three participants were excluded because of a computer error. Four additional participants were excluded because they failed to complete the experiment within the allotted time, resulting in a total of 55 participants. We compared the difference in mean judgments for time 1 and time 2 across problems and judgments. A difference (mean = -0.009, SD = .035) was detected between time 1 and time 2, t(135) = -3.10, p = .002. However, given the small magnitude of the difference, most of the variability in judgments appears to be noise rather than systematic.

Noise and errors

According to the PTV model, increased noise should be associated with increased errors and decreased semantic coherence. As in Experiment 1, error rates, semantic coherence
rates, and overall noise were computed for each subject. Similarly, we restricted our analyses for the minimum conditional errors to cases in which \( S(A) + S(B) > 1 \) and weighted the scores according to the number of observations. As shown in Table 2, the prediction held for conditional reversals and semantic coherence. However, the remaining correlations failed to reach statistical significance.

### Bayes’ theorem

We computed \( S(A)S(B|A) - S(B)S(A|B) \) for each participant’s judgments on each problem to test whether Bayes’ theorem holds stochastically. Although there was a systematic deviation, the difference was very close to zero, \( -0.008, SD = 0.12, 95\% CI [-0.012, -0.003] \), as predicted by the PTV model. Next, we tested whether the noise in the judgments is related to the variability in the distribution. For each set of judgments, we computed the absolute deviation and a composite measure of noise by summing the absolute differences in judgments at time 1 and time 2. Consistent with the model, the Spearman correlation was, \( r = 0.24, 95\% CI [0.21, 0.27] \).

### Model fit

The model was fit to aggregated data for each of the 34 problems following a similar procedure used in Experiment 1 (Appendix 1). The observed and predicted rates can be found in Table 5. Compared with Experiment 1, the absolute deviations were somewhat larger, mean = .09 (SD = .12), and the correlation between predicted and observed rates was somewhat lower, \( r = 0.38 \). Compared with overlapping and subset problems, the model performed relatively poorly on identical, independent, and mutually exclusive problems. Moreover, the model had difficulty accounting for semantic coherence, particularly for identical and independent sets, where the discrepancy was .65 and .30, respectively. Finally, the PTV model performed poorly on conversion errors. In general, the model predicts conditional reversals should be higher than conversion errors. However, the opposite trend was observed.

### DISCUSSION

By and large, the results of Experiment 2 mirrored those from Experiment 1. Some support was found for the PTV model’s critical property in which noise increases error rates. As predicted, increased noise was associated with more conditional reversals and less semantic coherence. The correlations between noise and the other errors were less conclusive because they were non-significant and somewhat underpowered for a medium effect size of \( r = 0.30 \) (power = .61, two-tailed). Consistent with the PTV model, judgments adhered stochastically to Bayes’ theorem. In addition, the noise in individual judgments was related to the absolute deviation from Bayes’ theorem. These results are consistent with the notion that noisy judgments were generated by a stochastic process that follows the rules of probability theory. However, two lines of evidence were at odds with the PTV model. First, conversion errors were generally higher than conditional reversals, a finding that suggests errors do not result exclusively from noise. This was reflected in the model’s tendency to underpredict conversion errors and overpredict conditional reversals. One possible explanation is that judgments are indeed noisy, but people may use a rule in which \( P(A|B) \) and \( P(B|A) \) are constrained to be equal (e.g., Reyna & Brainerd, 2008). Second, the PTV model performed poorly on semantic coherence for problems featuring independent and identical sets. This suggests that judgment noise alone is insufficient to account for the errors and semantic coherence. One reason is that semantic coherences for identical and independent sets are rare in the judgment space, making it unlikely that semantic coherence can be achieved through a purely random process. By contrast, the PTV model can account for semantic coherence on subsets and overlapping sets because they are relatively more common in the judgment space and lower semantic coherence rates are found empirically (Wolfe et al., 2013).

### GENERAL DISCUSSION

A common finding in the literature is that judgments depart from the norms of probability theory (e.g., Tversky & Kahneman, 1983; Barbey & Sloman, 2007). As a natural starting point, most psychological models attempt to account for non-normative judgments with assumptions that are inherently non-normative, such as denominator neglect in Fuzzy Trace Theory (Reyna & Brainerd, 2008) and weighted averaging in the CWA model (Nilsson et al., 2009). By contrast, the PTV model begins with the counter-intuitive assumption that judgments adhere to probability theory, but systematic errors result from noise in the judgments. In Experiment 1, we contrasted the PTV model with the CWA model. Many of the results were consistent with both models. For example, judgments showed stochastic adherence to the addition law and higher rates of double-conjunction, and

### Table 5. Predicted and observed error and semantic coherence rates for Experiment 2. Predicted rates are in parentheses

<table>
<thead>
<tr>
<th>Subset</th>
<th>MCE A</th>
<th>B</th>
<th>MCE B</th>
<th>A</th>
<th>CR</th>
<th>CE</th>
<th>SC</th>
<th>Mean Absolute Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identical</td>
<td>.04</td>
<td>.07</td>
<td>.03</td>
<td>.11</td>
<td>.03</td>
<td>.02</td>
<td>.06</td>
<td>.09</td>
</tr>
<tr>
<td>Independent</td>
<td>.17</td>
<td>.21</td>
<td>.17</td>
<td>.22</td>
<td>.11</td>
<td>.16</td>
<td>.23</td>
<td>.05</td>
</tr>
<tr>
<td>Mutually exclusive</td>
<td>.14</td>
<td>.24</td>
<td>.15</td>
<td>.24</td>
<td>.03</td>
<td>.14</td>
<td>.08</td>
<td>.02</td>
</tr>
<tr>
<td>Overlapping</td>
<td>.08</td>
<td>.12</td>
<td>.08</td>
<td>.11</td>
<td>.12</td>
<td>.16</td>
<td>.20</td>
<td>.07</td>
</tr>
<tr>
<td>Subset</td>
<td>.02</td>
<td>.15</td>
<td>.05</td>
<td>.12</td>
<td>.06</td>
<td>.07</td>
<td>.05</td>
<td>.02</td>
</tr>
</tbody>
</table>

MCE A|B, minimum conditional error for \( P(A|B) \); MCE B|A, minimum conditional error for \( P(B|A) \); CR, conditional reversal; CE, conversion error; SC, semantic coherence.
double-disjunctions fallacies were associated with more noise. The evidence provided more support for the PTV model in cases where the models made divergent predictions. For example, semantic coherence decreased as noise increased, a finding that was at odds with the CWA model. Also consistent with the PTV model, the increased noise in conjunctions due to order effects produced more conjunction fallacies. In Experiment 2, we tested similar predictions of the PTV model in conditional probability judgment. In line with the PTV model, semantic coherence decreased as noise increased and judgments adhered stochastically to Bayes’ theorem.

In several cases, however, the evidence was at odds with the PTV model. For example, we failed to find more noise in disjunctions compared to conjunctions, as implied by the integration rules of the PTV model. One could argue that this only qualifies as evidence against the one instantiation of the PTV model because there are alternative equations for conjunctions and disjunctions. However, the alternative equations should, by definition, produce the same predictions. For this reason, we argue that the failure to find more noise in disjunctions constitutes evidence against the model as a whole rather than a particular instantiation of the model (see Costello and Watts (2012) for a model with alternative assumptions). Another finding that was at odds with the PTV model was the higher rate of conversion errors compared with conditional reversals. This finding cannot be reconciled with simple random noise in judgment, except when the conditional probabilities are close to 1 (e.g., identical sets). In some cases, auxiliary assumptions could be modified to accommodate our findings. A response rule could be added to produce more conversion errors, or the assumption independence could be relaxed so that conjunctions are not necessarily noisier than disjunctions. However, such modifications would come at expense of computational tractability, parsimony, and by extension, falsifiability. Moreover, relaxing the assumption of independence may have unforeseen effects on the model’s ability to account for other phenomena in probability judgment, such as the conjunction fallacy.

Although the PTV model failed to provide a full account of our findings, it has been successful in advancing theory and illustrating the benefits of formal models. A major contribution of the PTV model and its predecessors (e.g., Erev et al., 1994) is the notion that systematic errors could, in principle, result from non-systematic sources of variability (i.e., noise). One benefit of considering the role of noise is that it may prevent theorists from needlessly positing complex mechanisms. In addition, several novel predictions regarding the variance in judgments were derived from the model, an aspect of judgment that is often neglected. Our results may have implications for the use of Bayesian models of cognition. Bayes’ theorem forms the basis of several models of cognition, ranging from perception to higher-order reasoning (Chater, Oaksford, Hahn, & Heit, 2010). The finding that judgments adhered stochastically to Bayes’ theorem provides some justification for using it as an approximation to some cognitive processes. At the same time, some caution should be exercised because not all aspects of the PTV model were supported and our findings on probability judgment may not generalize to some cognitive processes. Finally, the underlying premises of the PTV model add an interesting and complex dimension to debates about human rationality.

APPENDIX 1

In this section, we describe the model fitting procedure used to compute the predicted error and semantic coherence rates for joint and conditional probability judgment. As described in Costello (2009a), subjective probabilities are derived from a normal distribution that represents subjective confidence. Subjective probability distributions are produced via a response rule that maps the confidence distribution on the interval [0, 1] through a logistic function. To make the model more psychologically plausible, we assume that subjective probabilities are rounded to the nearest multiple of five, as commonly found in practice (Erev et al., 1994). Quantitative model predictions were derived by sampling from the subjective probability distributions for each of the 34 problems using group level data. Let $S(k)_{pdf}$ and $\sigma S(k)_{pdf}$ be the observed mean and standard deviation of event $k$ in problem $p$, formed by averaging judgments across participants at time 1 and time 2. To find the subjective probability distribution for event $k$ in problem $p$, $N=100,000$ simulated subjective confidence judgments were sampled from a normal distribution $C(k)_{pi} \sim N(\mu_{C(k)_{pi}}, \sigma C(k)_{pi})$. Next, the simulated subjective confidence judgments were transformed to subjective probabilities using the following logistic function: $S(k)_{pi} = \frac{1}{1 + e^{-\mu S A_{pi}}}$.

Parameters $\mu_{C(k)_{pi}}$ and $\sigma_{C(k)_{pi}}$ of the subjective confidence distribution were adjusted using the Nelder–Mead algorithm until the mean and standard deviation of the resulting subjective probability distribution, $S_{(k)_{pi}}$ and $\sigma S_{(k)_{pi}}$, approximated the observed mean and standard deviation as closely as possible.

Quantitative predictions for the error and semantic coherence rates were estimated by sampling $N=100,000$ times from the relevant estimated subjective probability distributions, $S_{(k)_{pdf}}$ and computing the relative frequency with which the errors and semantic coherence occurred. The predicted conjunction and disjunction fallacy rates were computed as $F_{pi} = S(B|A)_{pi} S(A)_{pi} - S(B)_{pi}$ and $P(CF)_{p} = P(DF)_{p} = \sum_{i=1}^{n} \left\{ \frac{f_{pi}}{N} \right\}$, where $CF$ and $DF$ denote the conjunction and disjunction fallacy, respectively. Similarly, the double-conjunction and double-disjunction fallacies were computed as $G_{pi} = S(A|B)_{pi} S(B)_{pi} - S(A)_{pi}$ and $P(DCF)_{p} = P(DDF)_{p} = \sum_{i=1}^{n} \left\{ \frac{g_{pi}}{N} \right\}$. The predicted minimum conjunction error was computed as $H_{pi} = -1 + S(A)_{pi} + S(B)_{pi} - S(A|B)_{pi} S(B)_{pi}$ and $P(MCE)_{p} = \sum_{i \neq j}^{n} \left\{ \frac{h_{pi}}{N} \right\}$, where $i \neq j$. The predicted maximum disjunction error was computed as $H_{pi} = \left[ S(A)_{pi} + S(B)_{pi} - S(B|A)_{pi} S(A)_{pi} \right] - \left[ S(A)_{pj} + S(B)_{pj} \right]$, and $P(MCE)_{p} = \sum_{i \neq j}^{n} \left\{ \frac{h_{pi}}{N} \right\}$. Semantic coherence was also estimated by computing the relative frequency with which it occurred in the simulated judgments (for details, refer to Wolfe & Reyna, 2010; Fisher & Wolfe, 2011). The model was fit by minimizing the sum of the squared differences between the predicted and observed error and
semantic coherence rates for each problem \( p \). \( \bar{S}(k)_p \) and \( \sigma(S(k)_p) \) were free to vary within ±0.02 of their observed values. This allowed the model to adjust for sampling error while being sufficiently constrained. One challenge in estimating the error and semantic coherence rates through simulation is that it destabilizes the parameter space; a slightly different rate will be obtained each time with the same parameters. Two measures were undertaken to minimize this problem. First, the predicted error rates were rounded to two decimal places. Second, a large number of simulated judgments (\( N = 100,000 \)) was used to further increase the stability of the estimates.

The model fitting procedure for conditional probability judgment in Experiment 2 was identical, except the rates were computed for the minimum conditional errors, the conditional reversal, the conversion error, and semantic coherence for conditional probability judgment. The predicted rate for the minimum conditional for \( A \cap B \) was, \( L_{pi} = S(A|\bar{A})_{pi} + S(B|\bar{B})_{pi} - S(\bar{A}|\bar{B})_{pi} \) and \( P(MCE A|B)_p = \sum_{i=0}^{1} P_{pi} \left( \frac{1}{1 - 0.1} \right) \). The minimum conditional error for \( B|A \) is computed similarly: \( M_{pi} = S(\bar{A}|\bar{B})_{pi} + S(\bar{B}|\bar{A})_{pi} - S(A|\bar{B})_{pi} \) and \( P(MCE B|A)_p = \sum_{i=0}^{1} P_{pi} \left( \frac{1}{1 - 0.1} \right) \). The conversion error is calculated as \( O_{pi} = 1 \) if \( S(A|\bar{B})_{pi} \neq S(\bar{B}|\bar{A})_{pi} \) and \( S(\bar{A}|\bar{B})_{pi} = S(\bar{B}|\bar{A})_{pi} \) and \( S(A|\bar{B})_{pi} \neq \bar{0} \) are true and \( O_{pi} = 0 \) otherwise. Thus, the probability of a conversion error is \( P(CE)_p = \sum_{i=0}^{1} P_{pi} \). A conditional reversal is coded as \( Q_{pi} = 1 \) if both \( S(A|\bar{B})_{pi} > S(\bar{B}|\bar{A})_{pi} \) and \( S(\bar{A}|\bar{B})_{pi} > S(B|\bar{A})_{pi} \) (\( O_{pi} = 0 \)) otherwise. Thus, the probability of a conditional reversal is \( P(CR)_p = \sum_{i=0}^{1} P_{pi} \).

REFERENCES


Authors’ biographies:

Christopher R. Wolfe. Ph.D. is a professor of Psychology at Miami University. His research interests include judgment and decision making and higher-order cognition.

Christopher R. Fisher is a doctoral candidate in Cognitive Psychology at Miami University. His research interests are probability judgment and decision making.

Authors’ addresses:

Christopher R. Fisher, Miami University, Oxford, OH, USA.

Christopher R. Wolfe, Miami University, Oxford, OH, USA.