

# On the provenance of judgments of conditional probability

Jiaying Zhao (jiayingz@princeton.edu)

Department of Psychology, Green Hall  
Princeton, NJ 08540 USA

Anuj Shah (akshah@princeton.edu)

Department of Psychology, Green Hall  
Princeton, NJ 08540 USA

Daniel Osherson (osherson@princeton.edu)

Department of Psychology, Green Hall  
Princeton, NJ 08540 USA

## Abstract

In standard treatments of probability,  $Pr(A|B)$  is defined as the ratio of  $Pr(A \cap B)$  to  $Pr(B)$ , provided that  $Pr(B) > 0$ . This account of conditional probability suggests a psychological question, namely, whether estimates of  $Pr(A|B)$  arise in the mind via implicit calculation of  $Pr(A \cap B)/Pr(B)$ . We tested this hypothesis (Experiment 1) by presenting brief visual scenes composed of forms, and collecting estimates of relevant probabilities. Direct estimates of conditional probability were not well predicted by  $Pr(A \cap B)/Pr(B)$ . Direct estimates were also closer to the objective probabilities defined by the stimuli, compared to estimates computed from the foregoing ratio. The hypothesis that  $Pr(A|B)$  arises from the ratio  $Pr(A \cap B)/[Pr(A \cap B) + Pr(\bar{A} \cap B)]$  fared better (Experiment 2).

**Keywords:** Conditional probability, reasoning, judgment;

## Introduction

Axiomatic presentations of probability (Ross, 1988) typically define conditional probability from absolute probability via the equation

$$(1) \quad Pr(A|B) =_{def} \frac{Pr(A \cap B)}{Pr(B)} \quad \text{provided } Pr(B) > 0.$$

As a theory of chance, the probability calculus can be justified independently of human psychology, for example, in terms of fair betting rates (Jeffrey, 2004). Definition (1) nonetheless invites the hypothesis that judgments of conditional probability arise by implicit calculation of the ratio of the two absolute probabilities shown above. Here we report experiments designed to test this hypothesis in a simple setting, in which probabilities must be estimated from brief visual presentation of forms of varying shape, color, and position.

The previous literature provides reasons to doubt that (1) reflects the provenance of conditional probability in the mind. According to (1),  $Pr(A|B) = Pr(B|A)$  only if  $Pr(A) = Pr(B)$ . Yet the inversion of conditional probabilities is a common feature of judgment even when it is recognized that  $Pr(A) \neq Pr(B)$  (Eddy, 1982; Dawes, Mirels, Gold, & Donahue, 1993). Such inversion undermines the conviction that most people understand the concept of conditional probability.

Another reason to doubt (1) is that it conflicts with intuition in cases involving continuous sample spaces (Hajek,

2003). For example, suppose a number is drawn uniform randomly from  $[0, 1]$ , and let  $B = \{.6, .7, .8\}$ .<sup>1</sup> It seems that the chance of falling below .75 given that a member of  $B$  is drawn equals  $2/3$  whereas (1) recognizes no such conditional probability because  $Pr(B) = 0$ . Examples of this character have prompted axiomatizations that reverse the roles of conditional and absolute probability. For example, conditional probability is primitive and  $Pr(A)$  is defined as  $Pr(A|\Omega)$  where  $\Omega$  is the certain event (Popper, 1959).

In what follows, we let RH denote the hypothesis that judgments of conditional probability arise from implicit calculation of the ratio shown in (1). To formulate predictions associated with RH, we rely on the following notation. Let two perceptual categories  $A, B$  be given (e.g., red, square), and let  $\mathbf{S}$  be the visual scene in question. Then, for a given experimental participant:

$Pr[\text{dir}](B)$  denotes the judged probability that a form drawn randomly from  $\mathbf{S}$  is  $B$ , and likewise for  $Pr[\text{dir}](A \cap B)$ .  
 $Pr[\text{dir}](A|B)$  denotes the judged probability that such a form is  $A$  assuming that it is  $B$ . (“dir” stands for “direct.”)

$Pr[\text{ind}](A|B)$  denotes the ratio of  $Pr[\text{dir}](A \cap B)$  to  $Pr[\text{dir}](B)$ . Thus,  $Pr[\text{ind}](A|B)$  is the conditional probability of  $A$  given  $B$  as computed from (1). (“ind” stands for “indirect.”)

$Pr[\text{obj}](A|B)$  denotes the percentage of  $A$ ’s in  $\mathbf{S}$  among the  $B$ ’s in  $\mathbf{S}$ , i.e., the true conditional probability in  $\mathbf{S}$  of  $A$  assuming  $B$  — and similarly for  $Pr[\text{obj}](B)$  and  $Pr[\text{obj}](A \cap B)$ . (“obj” stands for “objective.”)

We understand RH to entail:

1.  $Pr[\text{ind}](A|B)$  is an unbiased estimate of  $Pr[\text{dir}](A|B)$ .
2.  $Pr[\text{dir}](A|B)$  and  $Pr[\text{ind}](A|B)$  are equally close to  $Pr[\text{obj}](A|B)$ .

<sup>1</sup>The uniform distribution over  $[0, 1]$  sets the probability of sampling an interval  $I \subseteq [0, 1]$  equal to the length of  $I$ . Single points in  $[0, 1]$  thus have zero probability (since they represent intervals of length zero).

Another hypothesis will figure in the sequel. Since  $Pr(B) = Pr(A \cap B) + Pr(\bar{A} \cap B)$ , (1) implies:

$$(2) \quad Pr(A|B) = \frac{Pr(A \cap B)}{Pr(A \cap B) + Pr(\bar{A} \cap B)} \text{ provided } Pr(B) > 0.$$

Let RH' be the hypothesis that judgments of conditional probability arise from implicit calculation of the ratio shown in (2). Correspondingly, we let  $Pr[\text{ind}'](A|B)$  denote  $Pr[\text{dir}](A \cap B)$  divided by  $Pr[\text{dir}](A \cap B) + Pr[\text{dir}](\bar{A} \cap B)$ . The predictions of RH' are the same as for RH, with  $Pr[\text{ind}'](A|B)$  substituted for  $Pr[\text{ind}](A|B)$ .

We do not consider the hypothesis based on Bayes' Theorem

$$Pr(A|B) = \frac{Pr(B|A) \times Pr(A)}{Pr(B)}$$

inasmuch as conditional probability appears on both sides of the equation.

In the experiments reported below,  $Pr[\text{dir}](A|B)$  was elicited via two kinds of wording. In the *probability* condition, participants were asked a question of the form: "Suppose that a  $B$  is chosen at random from the array; what is the probability that it is an  $A$ ?" The *frequency* version of this question was: "What percent of the  $B$ 's in the array are  $A$ 's?" Similar wordings were used for  $Pr[\text{dir}](A \cap B)$  and  $Pr[\text{dir}](B)$ . The two formulations test the robustness of our results inasmuch as frequency formats sometimes yield estimates more consistent with the probability calculus (Tversky & Kahneman, 1983; Fiedler, 1988; Mellers, Hertwig, & Kahneman, 2001). In the present experiments, the impact of alternative formats was minimal.

## Experiment 1

The primary purpose of the first experiment was to test RH through its predictions 1 and 2.

### Participants

Forty-five undergraduate students from Princeton University participated in exchange for partial course credit (32 female, mean age 20.09 yrs, SD = 1.02).

### Materials

Participants viewed 12 sets of geometric shapes on a computer screen. Each set was a mixture of 20 triangles, squares, and circles in blue, red, and green (all three shapes and all three colors appeared in every matrix). A given set was shown four times, with each display lasting one second. The shapes in a given display were arrayed as a  $4 \times 5$  matrix, their respective positions individually randomized for each presentation. The purpose of multiple brief, randomized displays of a given set was to prevent responses based on counting.<sup>2</sup>

The four displays of a given set were initiated by a "Ready" button controlled by the participant. Henceforth, by a *trial*

<sup>2</sup>An alternative to this spatial display of stimuli is temporal presentation in which the shapes appear on the screen in a serial manner. This option is not explored in the current study.

associated with a given set is meant the successive display of its four randomized matrixes.

For each set we chose one color and one shape to serve as the categories  $A$  and  $B$  evoked in the Introduction. A different choice was made for each of the 12 sets; for six sets  $A$  was a color and  $B$  a shape, the reverse held for the other six. The sets were designed so that  $Pr[\text{obj}](A \cap B)$  and  $Pr[\text{obj}](B)$  were either .1 and .3, .4 and .6, or .8 and .9. These three cases yield  $Pr[\text{obj}](A|B)$  equal to .33, .67, or .89, respectively. Four sets fell into each of these cases, called *low*, *medium*, and *high levels* in what follows. Table 1 summarizes the objective probabilities figuring in the experiment.

Table 1: Objective probabilities in the sets of stimuli used in Experiments 1.

Level	$Pr[\text{obj}](A B)$	$Pr[\text{obj}](A \cap B)$	$Pr[\text{obj}](B)$
Low	0.33	0.1	0.3
Medium	0.67	0.4	0.6
High	0.89	0.8	0.9

### Procedure

Each participant served in both the *probability* and *frequency* conditions (the order was counterbalanced). In each condition, the participant viewed the 12 sets three times, once for each query  $Pr(B)$ ,  $Pr(A \cap B)$ , or  $Pr(A|B)$ . The colors and shapes representing  $A$  and  $B$  were the same in the three trials for a given set. The 36 resulting trials were presented in individualized random order under the constraint that a given set not appear twice in a row. Following each trial, the participant responded to one question corresponding to  $Pr(B)$ ,  $Pr(A \cap B)$ , or  $Pr(A|B)$ . For the probability condition, the questions are illustrated as follows.

#### SAMPLE PROBABILITY QUESTIONS:

$Pr(B)$	What is the probability that a randomly selected shape in the set is red?
$Pr(A \cap B)$	What is the probability that a randomly selected shape in the set is a red square?
$Pr(A B)$	What is the probability that a randomly selected shape in the set is square assuming that it is a red?

For the frequency condition, the corresponding questions were:

#### SAMPLE FREQUENCY QUESTIONS:

$Pr(B)$	What percent of the shapes in the set are blue?
$Pr(A \cap B)$	What percent of the shapes in the set are blue circles?
$Pr(A B)$	What percent of the blue shapes in the set are circles?

Thus, in both conditions, a given set yielded values for each of  $Pr[\text{dir}](B)$ ,  $Pr[\text{dir}](A \cap B)$  and  $Pr[\text{dir}](A|B)$ . Participants entered their answers using either decimals, fractions, or percents according to their preference. The experiment began with explanation of the task and practice trials. Participants were not informed that sets would be repeated (with different queries); none seem to have discovered this fact. Between the two conditions (probability and frequency), participants completed a 5-minute distraction task.

## Results

**Average responses** The probability and frequency conditions produced very similar numbers; across all participants, the average discrepancy between responses to corresponding queries was only 0.022. The two conditions were therefore collapsed.

For a given participant, we averaged the response to each query —  $Pr[\text{dir}](B)$ ,  $Pr[\text{dir}](A \cap B)$  or  $Pr[\text{dir}](A|B)$  — at each level (low, medium, high). Each of these nine categories of numbers (three levels by three queries) was then averaged across the 45 participants, yielding the results shown in Table 2 (standard deviations shown in parentheses). Comparison of Tables 1 and 2 (objective versus estimated probabilities) suggests that participants' judgments were reasonably accurate. In particular, the average value of  $Pr[\text{dir}](A|B)$  is close to  $Pr[\text{obj}](A|B)$  at all three levels (within 0.02, 0.04, and 0.07, respectively).

Table 2: Average estimates from Experiment 1.

Level	$Pr[\text{dir}](A B)$	$Pr[\text{dir}](A \cap B)$	$Pr[\text{dir}](B)$
Low	0.35 (0.10)	0.19 (0.08)	0.31 (0.07)
Medium	0.63 (0.10)	0.50 (0.10)	0.58 (0.07)
High	0.82 (0.12)	0.80 (0.07)	0.86 (0.04)

To compute  $Pr[\text{ind}](A|B)$ , for each participant and each level, we divided her average for  $Pr[\text{dir}](A \cap B)$  at that level by her average for  $Pr[\text{dir}](B)$ . Over the 45 participants, the means for  $Pr[\text{ind}](A|B)$  were 0.61 (SD = 0.21), 0.86 (SD = 0.17), and 0.94 (SD = 0.08) for the low, medium, and high levels, respectively.

**Test of RH** To test prediction 1, at each level we performed paired  $t$ -tests on the average values of  $Pr[\text{dir}](A|B)$  versus  $Pr[\text{ind}](A|B)$  across the 45 participants. In all three cases,  $Pr[\text{dir}](A|B)$  was reliably smaller than  $Pr[\text{ind}](A|B)$  (for low, medium, high levels, paired  $t(44) = 7.5, 7.7, 5.4$ , respectively,  $p < .01$ ). The differences between  $Pr[\text{dir}](A|B)$  and  $Pr[\text{ind}](A|B)$  are 0.26, 0.23, and 0.12 at the three levels.

To test prediction 2, at each level we computed for each participant the absolute difference between  $Pr[\text{dir}](A|B)$  and  $Pr[\text{obj}](A|B)$ , and between  $Pr[\text{ind}](A|B)$  and  $Pr[\text{obj}](A|B)$ . Across the 45 participants, the average absolute difference between  $Pr[\text{dir}](A|B)$  and  $Pr[\text{obj}](A|B)$  was 0.10, 0.10, and 0.08 at the three levels, compared to 0.30, 0.22, and 0.09 for

$Pr[\text{ind}](A|B)$ . Thus, direct estimates of objective conditional probability were more accurate than indirect at the low and medium levels (paired  $t$ -tests yield  $t(44) = 6.3$  and  $5.1$ ,  $p < .01$ ). The accuracy of direct estimates was close to that of indirect estimates at the high level ( $t(44) = 0.04$ ,  $p > .05$ ).

**Inversion of conditional probability** The data give scant evidence for confusion of  $Pr(B|A)$  with  $Pr(A|B)$ . For each participant at each level, we calculated the mean absolute difference between  $Pr[\text{obj}](A|B)$  and  $Pr[\text{dir}](A|B)$  along with the mean absolute difference between  $Pr[\text{obj}](B|A)$  and  $Pr[\text{dir}](A|B)$ . These means were based on the 8 estimates of  $Pr[\text{dir}](A|B)$  made at a given level. Inversion of the conditional would result in the absolute difference between  $Pr[\text{obj}](A|B)$  and  $Pr[\text{dir}](A|B)$  tending to be equal to or greater than the absolute difference between  $Pr[\text{obj}](B|A)$  and  $Pr[\text{dir}](A|B)$ . We found, however, that the first difference was smaller than the second at each level. The difference was significant at the low level (paired  $t(44) = 2.9$ ,  $p < .01$ ) but was just a trend at the medium and high levels (paired  $t(44) = 1.4$  in each case,  $p > .05$ ).

## Discussion of Experiment 1

The results of Experiment 1 are largely inconsistent with RH, the hypothesis that conditional probabilities are mentally calculated from the ratio appearing in Equation (1). The latter ratio consistently overestimated participants' direct judgments of  $Pr(A|B)$  (by more than 0.2 on average). Moreover, at the low and medium levels, direct estimates of objective conditional probability were considerably more accurate than estimates based on RH.

Finally, participants showed no sign of conflating  $Pr(A|B)$  with  $Pr(B|A)$ . Combined with the accuracy of their estimates of  $Pr[\text{obj}](A|B)$ , these results suggest they have a mature conception of conditional probability.

## Experiment 2

The second experiment was designed to replicate the first, and also to test Hypothesis RH', based on Equation (2). For this purpose, we added the query  $Pr(\bar{A} \cap B)$  to the three queries figuring in Experiment 1.

## Participants

Forty-five undergraduate students from Princeton University participated in exchange for partial course credit (35 female, mean age 19.2 yrs, SD = 1.35).

## Materials and Procedure

The stimuli from Experiment 1 were employed again. The procedure was the same except that each set figured in an additional trial that queried  $Pr(\bar{A} \cap B)$  as illustrated here.

PROBABILITY AND FREQUENCY QUERIES FOR  $Pr(\bar{A} \cap B)$ :

probability version: What is the probability that a randomly selected shape in the set is square and not red?

frequency version: What percent of the shapes in the set are square and not red?

Note that these queries have the form  $Pr(B \cap \bar{A})$  rather than the equivalent  $Pr(\bar{A} \cap B)$ . This was done to avoid ambiguity about the scope of the negation.  $Pr[\text{obj}](\bar{A} \cap B)$  was 0.2, 0.2, and 0.1 for low, medium, and high levels. To summarize, in both conditions a given set figured in four trials, one for each of the probabilities  $Pr[\text{dir}](B)$ ,  $Pr[\text{dir}](A \cap B)$ ,  $Pr[\text{dir}](\bar{A} \cap B)$  and  $Pr[\text{dir}](A|B)$ .

## Results

**Average responses** Once again, the probability and frequency conditions produced similar numbers; across all participants, the average discrepancy between responses to corresponding queries was only 0.024. The two conditions were therefore collapsed.

For a given participant, we averaged the response to each query —  $Pr[\text{dir}](B)$ ,  $Pr[\text{dir}](A \cap B)$ ,  $Pr[\text{dir}](\bar{A} \cap B)$  or  $Pr[\text{dir}](A|B)$  — at each level (low, medium, high). Each of these twelve categories of numbers (three levels by four queries) was then averaged across the 45 participants (see Table 3). In particular, the average value of  $Pr[\text{dir}](A|B)$  is close to  $Pr[\text{obj}](A|B)$  at all three levels (within 0.01, 0.04, and 0.06, respectively).

Table 3: Average direct estimates from Experiment 2.

Level	$B$	$A \cap B$	$\bar{A} \cap B$	$A B$
Low	0.29(0.05)	0.17(0.06)	0.23(0.07)	0.33(0.09)
Med	0.59(0.08)	0.51(0.16)	0.21(0.05)	0.63(0.10)
High	0.85(0.05)	0.78(0.06)	0.14(0.08)	0.83(0.06)

To compute  $Pr[\text{ind}'](A|B)$ , for each participant and each level, we divided her average for  $Pr[\text{dir}](A \cap B)$  at that level by her average for  $Pr[\text{dir}](A \cap B) + Pr[\text{dir}](\bar{A} \cap B)$ . Over the 45 participants, the means for  $Pr[\text{ind}'](A|B)$  were 0.44 (SD = 0.07), 0.71 (SD = 0.07), and 0.86 (SD = 0.07) for the low, medium, and high levels, respectively. The means for  $Pr[\text{ind}](A|B)$  were computed as in Experiment 1, yielding 0.60 (SD = 0.18), 0.87 (SD = 0.24), and 0.93 (SD = 0.09) at the three levels.

**Replication of Experiment 1** The results of Experiment 1 were replicated in Experiment 2. Regarding prediction 1 of Hypothesis RH, the average value of  $Pr[\text{dir}](A|B)$  across the 45 participants of Experiment 2 was reliably smaller than  $Pr[\text{ind}](A|B)$  at all three levels (paired  $t(44) = 8.9$ , 6.3, 5.9, respectively,  $p < .01$ ). The differences were 0.27, 0.24, and 0.10. Regarding prediction 2, direct estimates of  $Pr[\text{obj}](A|B)$  were closer than indirect estimates at all three levels ( $t(44) = 6.3$ , 3.9, 1.9,  $p < .01$  for the low and medium

levels,  $p < .05$  for the high level). Across the 45 participants, the average absolute difference between  $Pr[\text{dir}](A|B)$  and  $Pr[\text{obj}](A|B)$  was 0.09, 0.10, and 0.07 at the three levels, compared to 0.27, 0.24, and 0.10 for  $Pr[\text{ind}](A|B)$ .

Again, there was little evidence for conflation of  $Pr(B|A)$  with  $Pr(A|B)$ .  $Pr[\text{dir}](A|B)$  was significantly closer to  $Pr[\text{obj}](A|B)$  than to  $Pr[\text{obj}](B|A)$  at the low and high levels (paired  $t(44) = 2.8$ , 3.1, respectively,  $p < .01$ ), and also closer at the medium level but not significantly (paired  $t(44) = 1.1$ ,  $p > .05$ ).

**Test of RH'** To test the prediction that  $Pr[\text{ind}'](A|B)$  is an unbiased estimate of  $Pr[\text{dir}](A|B)$ , at each level we performed paired  $t$ -tests on the average values of  $Pr[\text{dir}](A|B)$  versus  $Pr[\text{ind}'](A|B)$  across the 45 participants. For all three levels,  $Pr[\text{dir}](A|B)$  was reliably smaller than  $Pr[\text{ind}'](A|B)$  ( $t(44) = 6.3$ , 4.3, 2.3, respectively,  $p < .01$ ), with gaps of 0.11, 0.08, and 0.03. Although  $Pr[\text{ind}'](A|B)$  systematically overestimates  $Pr[\text{dir}](A|B)$ , its gaps are smaller than for  $Pr[\text{ind}](A|B)$ . RH' thus appears to be more accurate than RH.

The superiority of RH' must be due to the denominator  $Pr(A \cap B) + Pr(\bar{A} \cap B)$  in (2) compared to  $Pr(B)$  in (1) inasmuch as the respective numerators are identical. Indeed,  $Pr[\text{dir}](A \cap B) + Pr[\text{dir}](\bar{A} \cap B)$  typically exceeded  $Pr[\text{dir}](B)$  despite their equivalence in the probability calculus. Across the 12 stimuli and 45 participants, the mean of  $Pr[\text{dir}](A \cap B) + Pr[\text{dir}](\bar{A} \cap B)$  is 0.68 (SD = 0.08) whereas the mean of  $Pr[\text{dir}](B)$  is 0.58 (SD = 0.04). The slightly greater value of the denominator in (2) lowers the value of the ratio thereby mitigating the overestimation of  $Pr[\text{dir}](A|B)$ .

To test the prediction that  $Pr[\text{dir}](A|B)$  and  $Pr[\text{ind}'](A|B)$  are equally close to  $Pr[\text{obj}](A|B)$ , at each level we computed for each participant the absolute difference between  $Pr[\text{dir}](A|B)$  and  $Pr[\text{obj}](A|B)$ , and between  $Pr[\text{ind}'](A|B)$  and  $Pr[\text{obj}](A|B)$ .

At the low level,  $Pr[\text{dir}](A|B)$  was significantly more accurate than  $Pr[\text{ind}'](A|B)$  ( $t(44) = 2.0$ ,  $p < .05$ ) whereas  $Pr[\text{ind}'](A|B)$  was more accurate than  $Pr[\text{dir}](A|B)$  at the other levels; for the medium level the difference was significant ( $t(44) = 2.0$ ,  $p < .05$ ) but just a trend at the high level ( $t(44) = 0.8$ ,  $p > .05$ ). Across the 45 participants, the average absolute difference between  $Pr[\text{dir}](A|B)$  and  $Pr[\text{obj}](A|B)$  was 0.09, 0.10, and 0.07 at the three levels, compared to 0.12, 0.08, and 0.06 for  $Pr[\text{ind}'](A|B)$ .

As a predictor of  $Pr[\text{obj}](A|B)$ ,  $Pr[\text{ind}'](A|B)$  was superior to  $Pr[\text{ind}](A|B)$ . In fact, 35 out of the 45 participants showed smaller average, absolute deviation between  $Pr[\text{obj}](A|B)$  and  $Pr[\text{ind}'](A|B)$  than between  $Pr[\text{obj}](A|B)$  and  $Pr[\text{ind}](A|B)$  ( $p < 0.05$ ).

## Discussion of Experiment 2

The results of Experiment 1 were replicated in the present study.  $Pr[\text{ind}](A|B)$  markedly overestimated  $Pr[\text{dir}](A|B)$ , and also predicted  $Pr[\text{obj}](A|B)$  less well than  $Pr[\text{dir}](A|B)$ . Also,  $Pr[\text{dir}](A|B)$  was closer to  $Pr[\text{obj}](A|B)$  than to  $Pr[\text{obj}](B|A)$ , providing no evidence for systematic conflation.

tion of conditionals with their inverse.

The novel finding is the greater accuracy of  $Pr[\text{ind}'](A|B)$  compared to  $Pr[\text{ind}](A|B)$  at predicting  $Pr[\text{dir}](A|B)$ . Although  $Pr[\text{ind}'](A|B)$  overestimates  $Pr[\text{dir}](A|B)$ , its error is about half that of  $Pr[\text{ind}](A|B)$ . Moreover,  $Pr[\text{ind}'](A|B)$  predicts  $Pr[\text{obj}](A|B)$  about as well as does  $Pr[\text{dir}](A|B)$ .  $\text{RH}'$  is thus better supported than  $\text{RH}$  by our data.

## General Discussion

Our findings suggest that judgments of conditional probability do not arise from mental division of the kind envisioned in the standard definition. For, the ratio  $Pr(A \cap B)/Pr(B)$  seen in Equation (1) systematically overestimates such judgments in all three of our experiments. Compared to direct estimates, the ratio is also further from the objective conditional probabilities inherent in the stimuli, providing another perspective on the limitations of (1) as a psychological theory. The less familiar but equivalent definition (2) comes closer to predicting raw judgments but it also errs on the side of overestimation. At the same time, (2) was as accurate as raw judgment in predicting objective conditional probabilities.

The superiority of  $\text{RH}'$  to  $\text{RH}$  is due to the slight overestimation of  $Pr(B)$  when it is decomposed as  $Pr(A \cap B) + Pr(\bar{A} \cap B)$ . Such decomposition often (but not invariably) increases estimates of event probability (Tversky & Koehler, 1994; Sloman, Rottenstreich, Wisniewski, Hadjichristidis, & Fox, 2004). Overall, neither  $\text{RH}$  nor  $\text{RH}'$  seems accurate as an account of the provenance of conditional probability.

Finally, in neither Experiment 1 nor 2 did we find evidence for conflation of  $Pr(A|B)$  with  $Pr(B|A)$ . At least in this respect, our participants seem to have understood the questions they were posed. Indeed, as seen earlier,  $Pr[\text{dir}](A|B)$  was impressively close to  $Pr[\text{obj}](A|B)$ .

The sample space in our experiments is transparent, and all probabilities were grounded in frequencies. The results thus discredit  $\text{RH}$  and  $\text{RH}'$  when probability is *extensional*. In this setting it is easy to envision theories of conditional probability that are alternative to the ratio accounts (1) and (2). Asked about  $Pr(\text{red}|\text{square})$ , for example, one might attempt to focus attention on just the squares then estimate the proportion of reds in this set. When the underlying partition of events is less evident than here it may be challenging to identify the relevant symmetries, opening the door to misconceptions and biases (Fox & Levav, 2004). Event-counting seems nonetheless central to many extensional settings, in which probability can be defined from frequency.

Matters are more complicated for the *intensional case*, which involves probabilities of non-repeatable events, for example:

$B$  = NASA merges with the European Space Agency by 2030.

$A$  = Humans walk on Mars by 2050.

It is unclear how a counting scheme could be deployed to construct  $Pr(A|B)$ . Ratio hypotheses like  $\text{RH}$  and  $\text{RH}'$  are thus all the more attractive for such events. Moreover, the proviso that

$Pr(A|B)$  is defined only if  $Pr(B) > 0$  seems more palatable in the intensional setting compared to the counterintuitive results it produces extensionally (Hajek, 2003). Unfortunately, ratio hypotheses rely on  $Pr(A \cap B)$ , and how this quantity is mentally calculated appears to be just as mysterious in the intensional framework as the calculation of  $Pr(A|B)$ . Both require determining the compatibility of  $A$  and  $B$ .

In any event, the poor performance of  $\text{RH}$  and (to a lesser extent)  $\text{RH}'$  in predicting judgments of conditional probability within our extensional framework suggests that they are unsatisfactory as well for the intensional case. But only further research will decide the matter.

## Acknowledgments

Osherson acknowledges support from the Henry Luce Foundation.

## References

Dawes, R., Mirels, H. L., Gold, E., & Donahue, E. (1993). Equating inverse probabilities in implicit personality judgments. *Psychological Science*, 4(6), 396–400.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, 50, 123–129.

Fox, C. R., & Levav, J. (2004). Partitioneditcount: Naive extensional reasoning in judgment of conditional probability. *Journal of Experimental Psychology: General*, 133(4), 626–642.

Hajek, A. (2003). What conditional probability could not be. *Synthese*, 137, 273–323.

Jeffrey, R. C. (2004). *Subjective probability: The real thing*. Cambridge: Cambridge University Press.

Mellers, B. A., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? an exercise in adversarial collaboration. *Psychological Science*, 12, 269–275.

Popper, K. (1959). *The logic of scientific discovery*. New York: Basic Books.

Ross, S. (1988). *A first course in probability*, 3rd edition. New York City: Macmillan.

Sloman, S. A., Rottenstreich, Y., Wisniewski, E., Hadjichristidis, C., & Fox, C. R. (2004). Typical versus atypical unpacking and superadditive probability judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 573–582.

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

Tversky, A., & Koehler, D. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, 101(4), 547–567.