

# Descriptive Assessment of Jeffrey's Rule

Jiaying Zhao (jiayingz@princeton.edu)

Department of Psychology, Green Hall,  
Princeton University, NJ 08540 USA

Daniel Osherson (osherson@princeton.edu)

Department of Psychology, Green Hall,  
Princeton University, NJ 08540 USA

## Abstract

Jeffrey (1983) proposed a generalization of conditioning as a means of updating probability distributions when new evidence drives no event to certainty. His rule requires the stability of certain conditional probabilities through time. We tested this assumption ("invariance") from the psychological point of view. In Experiment 1 participants offered probability estimates for events in Jeffrey's candlelight example. Two further scenarios were investigated in Experiment 2, one in which invariance seems justified, the other in which it does not. Results were in rough conformity to Jeffrey (1983)'s principle.

**Keywords:** Jeffrey's rule; invariance; probability updating;

## Introduction

Consider an idealized agent whose beliefs are represented by a (subjective) probability distribution  $Pr_1$  over an outcome space  $\mathcal{S}$ . Let  $B \subseteq \mathcal{S}$  be such that  $Pr_1(B) > 0$  and suppose that experience intervenes to convince the agent that  $B$  is certainly true. What probability distribution  $Pr_2$  should represent the agent's new state of belief? The Bayesian answer (Hacking, 2001, Ch. 15) identifies  $Pr_2$  with the result of conditioning  $Pr_1$  on  $B$ , that is,  $Pr_2(\cdot) = Pr_1(\cdot|B)$ . Much can be said in favor of the latter equation from the normative perspective. For example, it follows from compelling axioms on belief change (Gardenfors, 1988, §5.2), and its violation exposes the agent to sure-loss betting contracts (Harman, 1999, §4.12). Updating has also been examined from the psychological perspective with focus on the use of Bayes' Theorem to compute conditional probability (see Stanovich (2010, Ch. 3)).

Conditioning is not always suited, however, to represent the impact of new information. In particular, Jeffrey (1983, §11.1) notes that the passage of experience need not raise the probability of any event to one. He gives the example of examining cloth by faint candlelight. The cloth's potential colors might correspond to different events over  $\mathcal{S}$  but none may become certain as a result of the examination. Nor is it feasible to augment  $\mathcal{S}$  to include visual sensations, with the idea of setting one of them to unity. Such sensations are too difficult to express and individuate. Instead, says Jeffrey, "the best we can do is to describe, not the quality of the visual experience itself, but rather its effects on the observer," for example, that the probability of blue has shifted to .75 from its original value.

To fill in the rest of  $Pr_2$  after experience has set the value of  $Pr_2(B)$ , Jeffrey relies on the law of total probability. Let  $G \subseteq \mathcal{S}$  be given, and suppose that  $0 < Pr_2(B) < 1$ . Then:

$$(1) \quad Pr_2(G) = Pr_2(G|B)Pr_2(B) + Pr_2(G|\bar{B})Pr_2(\bar{B}).$$

If experience has not influenced the conditional probability of  $G$  given  $B$  nor that of  $G$  given  $\bar{B}$  then *invariance* is said to hold (Jeffrey, 2004, §3.2). That is:

$$(2) \quad Pr_2(G|B) = Pr_1(G|B) \quad \text{and} \quad Pr_2(G|\bar{B}) = Pr_1(G|\bar{B}).$$

Substituting (2) into (1) yields:

$$(3) \quad Pr_2(G) = Pr_1(G|B)Pr_2(B) + Pr_1(G|\bar{B})Pr_2(\bar{B}).$$

(3) is known as "Jeffrey's rule." It shows how change in the probability of  $B$  is propagated to  $G$ , without experience directly affecting  $G$ . It is straightforward to generalize (3) to finer partitions, in place of the binary partition  $B, \bar{B}$ .

Assuming invariance, it is easy to show that (3) defines a genuine probability distribution  $Pr_2$ , and in the special case of  $Pr_2(B) = 1$ , that it agrees with conditionalization. Moreover, Williams (1980) proves that  $Pr_2$  as given by (3) is the closest distribution to  $Pr_1$  that yields the new probability of  $B$ , where "closeness" is measured by cross-entropy with respect to  $Pr_1$ . The normative status of Jeffrey's rule has nonetheless been questioned because successive uses produce distinct distributions depending on the order in which events are considered (Doring, 1999). In our view, such doubts disappear upon closer inspection of the evidential weight of probability judgments (Wagner, 2002; Osherson, 2002).

It is important to observe that Invariance (2) is not normatively justified in every situation. Sometimes the conditional probabilities are shifted by experience. For example, let  $G$  represent vigorous growth of a potted plant, and let  $B$  represent the decision to place it in the bedroom. Then noticing ample sunshine in the bedroom would increase not just the probability of  $B$  but also the probability of  $G$  given  $B$ . In contrast, if all you notice is the absence of plants in the bedroom then the probability of  $B$  increases without a change in the probability of  $G$  given  $B$ , yielding invariance.

To decide whether invariance is warranted in a given situation, we rely on an observation due to Pearl (1988). Given an experience  $e$  that intervenes between times 1 and 2, we expect invariance to hold if at time 1,  $G$  is conditionally independent of  $e$  given  $B$ , that is:

$$(4) \quad \text{Conditional independence of } G \text{ from } e$$

given  $B$ :  $Pr_1(G | B, e) = Pr_1(G | B)$ .

To see that invariance depends on (4), observe that  $Pr_2(GB) = Pr_1(GB, e)$  since  $e$  is what transpires between times 1 and 2, and the latter term equals  $Pr_1(GB)$  just in case (4) holds.

The focus of the present paper is whether invariance is descriptively accurate when mandated normatively. In Experiment 1 undergraduates offer probabilities for events in Jeffrey's candle example. Of particular interest is the extent to which invariance is honored. Two further scenarios are then investigated, one in which invariance seems justified, the other in which it does not. The results show modest deviations from invariance where it seems justified normatively.

## Experiment 1

### Participants

Ninety-six undergraduates from Princeton University participated in exchange for partial course credit (41 female, mean age 19.4 yrs,  $SD = 1.0$ ).

### Materials

We simulated Jeffrey's candle example by having participants examine colored paper cards with a dim flashlight. There were 12 blue cards and 38 purple cards. Each card was marked with either a hippopotamus or a giraffe on one side. Of the blue cards, eight were marked with a hippo and four with a giraffe. Of the purple cards, 24 were marked with a hippo and 14 with a giraffe. We chose *giraffe* and *blue* as the categories  $G$  and  $B$  evoked in the Introduction. Table 1 summarizes the objective probabilities figuring in the experiment.

### Procedure

Sixty-four participants performed the *experimental* condition, and 32 performed the *control* condition. The purpose of the control condition was to assess the impact of being asked to evaluate the same probabilities a second time. In the experimental condition, the experimenter first shuffled the cards and showed each to the participant. Then the experimenter turned away from the participant, drew one card from the shuffled deck, and put it in her pocket. The draw appeared to be random but in fact was guaranteed across participants to deliver equal numbers of blue and purple cards (for statistical purposes). The participant was informed that the card was randomly chosen, and then answered the following questions about the card, via computer interface (the order was randomized for each participant):

#### PROBABILITY QUESTIONS:

$Pr(G)$	What's the probability that there is a giraffe on the card?
$Pr(B)$	What's the probability that the card is blue?
$Pr(G B)$	What's the probability that there is a giraffe on the card assuming that the card is blue?
$Pr(G \bar{B})$	What's the probability that there is a giraffe on the card assuming that the card is purple?
$Pr(B G)$	What's the probability that the card is blue assuming that there is a giraffe on the card?
$Pr(B \bar{G})$	What's the probability that the card is blue assuming that there is a hippo on the card?

The estimates  $Pr(B|G)$  and  $Pr(B|\bar{G})$  served as a contrast with  $Pr(G|B)$  and  $Pr(G|\bar{B})$ . Given our procedure (see below), the former estimates were not expected to be invariant across the flashlight experience whereas the latter were.

The participant was then informed that s/he would briefly see the card under dim light. The card would be placed face down on the table so that the participant would not see the animal but only the color of the card. The experimenter then turned off the lights in the room, moved the card from her pocket to the table, and flashed the light for about one second. The card was then returned to the experimenter's pocket, and the participant answered the same set of questions shown above (in a different random order). Since participants had to give their estimates twice to the same questions, we informed them that they were free to provide the same estimate or a different estimate the second time around.

In the control condition, the procedure was the same except that the light was applied to the chosen card immediately after its draw. Participants in the control condition thus answered the questions shown above just once, after briefly seeing the card under the dim light.

### Results

**Average responses.** We separately analyzed results for participants exposed to blue cards (Blue group) and those exposed to purple cards (Purple group). In the experimental condition we use  $Pr_1$  to refer to probability estimates before the light experience and  $Pr_2$  for estimates after the experience. For each condition and each color group, we averaged  $Pr_1$  and  $Pr_2$  estimates. The results are shown in Table 1.

Table 1: Average probably estimates by participants and objective probabilities in Experiment 1 (standard deviations in parentheses).

	$Pr(G)$	$Pr(B)$	$Pr(G B)$
Blue $Pr_1$	0.36(0.12)	0.33(0.11)	0.36(0.16)
Blue $Pr_2$	0.38(0.15)	0.71(0.26)	0.37(0.19)
Purple $Pr_1$	0.37(0.12)	0.31(0.12)	0.34(0.15)
Purple $Pr_2$	0.37(0.12)	0.17(0.16)	0.28(0.20)
Blue control	0.37(0.18)	0.70(0.26)	0.39(0.16)
Purple control	0.35(0.09)	0.22(0.11)	0.35(0.12)
Objective	0.36	0.24	0.33
	$Pr(G \bar{B})$	$Pr(B G)$	$Pr(B \bar{G})$
Blue $Pr_1$	0.34(0.16)	0.39(0.20)	0.36(0.14)
Blue $Pr_2$	0.39(0.16)	0.51(0.27)	0.53(0.25)
Purple $Pr_1$	0.36(0.14)	0.34(0.16)	0.34(0.17)
Purple $Pr_2$	0.36(0.14)	0.23(0.25)	0.27(0.26)
Blue control	0.30(0.16)	0.58(0.26)	0.48(0.27)
Purple control	0.31(0.11)	0.32(0.20)	0.24(0.15)
Objective	0.37	0.22	0.25

**Control vs. experimental conditions.** As experimental participants provided estimates twice to the same questions,  $Pr_2$  estimates were compared to those of control group. Independent-sample  $t$ -tests for each of the six questions revealed no reliable differences between  $Pr_2$  and the control estimates. Thus, these participants seem not to have been influenced by having to evaluate the same probabilities twice.

**Analysis of the Experimental Blue group.** We report the Experimental Blue and Experimental Purple participants separately, starting with Blue. As a manipulation check, we first determined whether  $Pr(B)$  increased after participants saw the blue card under dim light. As expected,  $Pr_2(B)$  was reliably larger than  $Pr_1(B)$  [paired  $t(31) = 7.0, p < .01$ , Wilcoxon  $p < .01$ ].

To see whether invariance holds as described in (2), we compared  $Pr_1(G|B)$  to  $Pr_2(G|B)$  via paired  $t$ -test and found no reliable difference [ $df = 31, p > .05$ ]. Of the 32 Blue participants, 18 offered a different  $Pr_2(G|B)$  estimate from  $Pr_1(G|B)$ . For these 18 participants, the average signed difference between  $Pr_1(G|B)$  and  $Pr_2(G|B)$  is  $-0.02$  which is not reliably different from 0 [ $t(17) = 0.3, p > .05$ ].  $Pr_1(G|\bar{B})$  was likewise found to be close to  $Pr_2(G|\bar{B})$  [paired  $t(31) = 1.3, p > .05$ ]. Eighteen out of 32 participants gave a different  $Pr_2(G|\bar{B})$  estimate from  $Pr_1(G|\bar{B})$ . The average signed difference between  $Pr_1(G|\bar{B})$  and  $Pr_2(G|\bar{B})$  is  $-0.09$  which is reliably different from 0 [ $t(17) = 2.2, p < .05$ ].

To more precisely quantify violation of invariance, for each participant we calculated the absolute movement between two estimates as a percentage of the original estimate, via:

$$(5) \quad \text{Invariance violation} = \frac{|Pr_2(G|B) - Pr_1(G|B)|}{Pr_1(G|B)}$$

We compared invariance violations to the movement of the converse probability (i.e., blue given giraffe), computed via:

$$(6) \quad \text{Converse movement} = \frac{|Pr_2(B|G) - Pr_1(B|G)|}{Pr_1(B|G)}$$

Following an analysis due to Pearl (1988), we expected the converse movements to exceed the invariance violations. This is because giraffe seems to be conditionally independent of the light-experience given blue (once you know that the card is blue, the light provides no further information) whereas blue seems not to be conditionally independent of the light-experience given giraffe (the light here provides additional information about the color). Consistent with this expectation, the means for invariance violations and converse movements were 37.6% and 88.3%, respectively. This difference is reliable by paired  $t$ -test ( $p < .05$ ) and Wilcoxon test ( $p < .01$ ).

For each participant we also computed invariance violation for  $Pr(G|\bar{B})$  via the following:

$$(7) \quad \text{Invariance violation for } \bar{B} = \frac{|Pr_2(G|\bar{B}) - Pr_1(G|\bar{B})|}{Pr_1(G|\bar{B})}$$

The mean invariance violation of  $Pr(G|\bar{B})$  was 45.9% and was not reliably different from the 37.6% violation of  $Pr(G|B)$  reported above [paired  $t(31) = 0.6, p > .05$ ].

From the results above, invariance seems to hold at least approximately. We therefore asked about its use in updating the probability of  $G$ . Specifically, for each participant, we computed the value of  $Pr(G)$  via the law of total probability (1), relying on the participant's estimates

for the quantities at the right of the equality. We will call this value the *total* probability of  $G$ , or  $Pr_{total}(G)$  for short. Likewise, for each participant we computed the value of  $Pr(G)$  via Jeffrey's rule (3). We will call this value the *Jeffrey* probability of  $G$ , or  $Pr_{Jeff}(G)$  for short. The latter estimates were compared to the participant's direct evaluation of  $Pr_2(G)$  via absolute difference:

$$(8) \quad \begin{aligned} \text{total error} &= |Pr_2(G) - Pr_{total}(G)| \\ \text{Jeffrey error} &= |Pr_2(G) - Pr_{Jeff}(G)| \end{aligned}$$

The means for total and Jeffrey error were .07 and .10, respectively, not reliably different via paired  $t$ -test [ $t(31) = 1.7, p > .05$ ] or Wilcoxon test ( $p > .05$ ).

**Analysis of the Experimental Purple group.** We first determined whether  $Pr(B)$  decreased after participants saw the purple card under dim light. As expected,  $Pr_2(B)$  was reliably less than  $Pr_1(B)$  [paired  $t(31) = 5.6, p < .01$ , Wilcoxon  $p < .01$ ].

We compared  $Pr_1(G|B)$  to  $Pr_2(G|B)$  via paired  $t$ -test and found no reliable difference [ $df = 31, p > .05$ ]. Of the 32 Purple participants, 21 offered a different  $Pr_2(G|B)$  estimate from  $Pr_1(G|B)$ . The average signed difference between  $Pr_1(G|B)$  and  $Pr_2(G|B)$  was 0.10 which was not reliably different from 0 [ $t(20) = 1.8, p > .05$ ].  $Pr_1(G|\bar{B})$  was also found to be close to  $Pr_2(G|\bar{B})$  [paired  $t(31) = 1.6, p > .05$ ]. The average signed difference between  $Pr_1(G|\bar{B})$  and  $Pr_2(G|\bar{B})$  was  $-0.01$  which was not reliably different from 0 [ $t(15) = 0.3, p > .05$ ].

Just as for the Blue group, we computed invariance violations of  $Pr(G|B)$  using (5) and converse movement of  $Pr(B|G)$  using (6). The means for invariance violations and converse movements were 48.6% and 60.3%, respectively. This difference is reliable by paired  $t$ -test ( $p < .05$ ) and Wilcoxon test ( $p < .05$ ). The mean invariance violation for  $Pr(G|\bar{B})$  via (7) was 18.9%, reliably smaller than the 48.6% violation of  $Pr(G|B)$  [paired  $t(31) = 2.9, p < .01$ ].

Once again, invariance seems to hold at least approximately so for each participant, we computed the value of  $Pr(G)$  via the law of total probability (1), again denoting this value by  $Pr_{total}(G)$ . Likewise, for each participant we computed the value of  $Pr(G)$  via Jeffrey's rule (3), denoting this value by  $Pr_{Jeff}(G)$ . The latter estimates were compared to  $Pr_2(G)$  via the absolute differences shown in (8). The means for total and Jeffrey error were .05 and .07, respectively, not reliably different via paired  $t$ -test [ $t(31) = 1.5, p > .05$ ] or Wilcoxon test ( $p > .05$ ).

## Discussion of Experiment 1

In the procedure of Experiment 1, respect for the invariance principle (2) seems normatively mandated inasmuch as experience with the light provides no further information about  $G$  once it is granted that the card is blue. In other words,  $Pr_2(G|B) = Pr_1(G|B, \ell) = Pr_1(G|B)$ , where  $\ell$  is the experience provided by the light (as discussed in the Introduction).

A majority of participants, in contrast, gave different estimates for  $Pr_2(G|B)$  compared to  $Pr_1(G|B)$  after gaining new information about color via the light.

As a percentage of  $Pr_1(G|B)$ , the absolute difference between  $Pr_2(G|B)$  and  $Pr_1(G|B)$  was not trivial but nonetheless reliably smaller than the absolute percentage difference for the converse probabilities  $Pr_2(B|G)$  and  $Pr_1(B|G)$ . Normatively, invariance is not expected with respect to  $Pr(B|G)$ . When used to estimate  $Pr_2(G)$  via the law of total probability, we saw that  $Pr_1(G|B)$  could be substituted for  $Pr_2(G|B)$  with little loss of accuracy [total versus Jeffrey error, as in (8)]. This provides another indication of the relative modesty of invariance violations in Experiment 1.

## Experiment 2

For another assessment of invariance, we asked a new group of participants to estimate probabilities for events in two different scenarios. The *lottery* scenario was designed to justify invariance whereas the *ultimatum game* scenario was not.

### Participants

One hundred undergraduates from Princeton University participated in exchange for partial course credit (58 female, mean age 19.5 yrs, SD = 1.1). None had served in Experiment 1.

### Materials and procedure

Fifty participants served in the *experimental* condition, and another 50 in the *control* condition. As in Experiment 1, the purpose of the control condition was to assess the impact of being asked to evaluate the same probabilities a second time. Each participant in both conditions was presented with the lottery and the ultimatum game scenarios (the order was counterbalanced).

**Lottery scenario.** In this scenario we substitute  $C$  (*buying a car*) for  $G$  and  $W$  (*winning the lottery*) for  $B$ . The following description was presented on the computer screen for each participant:

Imagine that a randomly chosen adult (call him Mr. X) in New Jersey has just purchased the *Jersey Cash 5* lottery for this week. In this lottery, there are 5 numbers to be drawn, each from 1 to 40. Each number is drawn from the bowl and then put aside. The lottery jackpot is \$240,000 which will be shared by players who have all 5 winning numbers (the order of the numbers doesn't matter). The numbers on Mr. X's lottery ticket are 12 17 24 32 39.

In the experimental condition, the participant answered the following questions before the lottery numbers were drawn (the order was randomized for each participant):

### LOTTERY PROBABILITY QUESTIONS:

$Pr(C)$	What's the probability that Mr. X will buy a new car in the next two years?
$Pr(W)$	What's the probability that Mr. X will win the jackpot?
$Pr(C W)$	What's the probability that Mr. X will buy a new car in the next two years assuming that he wins the jackpot?
$Pr(C \bar{W})$	What's the probability that Mr. X will buy a new car in the next two years assuming that he does NOT win the jackpot?

The participant was then presented with the following additional information.

It's the night of the lottery, and the numbers are being drawn. Mr. X becomes excited because the first four draws are 32, 12, 24, and 17. In other words, the first four numbers drawn match the numbers on his ticket.

The participant answered the same set of questions shown above (in a different random order) before the last number was drawn. Since participants had to give their estimates twice to the same questions, we informed them that they were free to provide the same estimate or a different estimate the second time around. In the control condition, participants saw the description of the lottery immediately followed by the results of the first four numbers, and then answered the questions shown above just once.

In the lottery scenario knowing the results of the first four draws provides no further information about  $W$  once it is granted that Mr. X wins the lottery. In other words,  $Pr_2(C|W) = Pr_1(C|W, f) = Pr_1(C|W)$ , where  $f$  is the experience of knowing the results of the first four draws. For this reason, invariance seems justified.

**Ultimatum game scenario.** Here we use  $A$  (*accepting the offer*) and  $O$  (*offering at least \$4*) to replace  $G$  and  $B$ . The following description was presented on the computer screen for each participant:

Imagine that two undergraduate students are randomly chosen from Princeton University to play a game. The game works as follows. The two students are given the opportunity to split \$10. One student is the proposer and the other is the responder. The proposer makes an offer as to how \$10 should be split between the two. The responder can either accept or reject this offer. If the responder accepts the offer, the money is split as proposed, but if the responder rejects the offer, then neither of them receives anything. The students have just finished the first trial of the game.

In the experimental condition, the participant was informed that the two students were about to play the second trial. The participant then answered the following questions (randomly ordered) about the second trial *prior to learning the outcome of the first trial*.

#### ULTIMATUM GAME PROBABILITY QUESTIONS:

$Pr(A)$	What's the probability that the responder will accept the offer from the proposer in the second trial?
$Pr(O)$	What's the probability that the proposer will offer AT LEAST \$4 to the responder in the second trial?
$Pr(A O)$	What's the probability that the responder will accept the offer assuming that the proposer offers AT LEAST \$4 in the second trial?
$Pr(A \bar{O})$	What's the probability that the responder will accept the offer assuming that the proposer offers LESS THAN \$4 in the second trial?

The participant was then presented with the following additional information about the scenario:

Now you learn that in the first trial the responder rejected the proposer's offer and neither of them received anything. They are about to play the second trial.

The participant answered the same set of questions shown above (in a different random order) about the second trial of the game. Again we informed participants that they were free to provide the same estimate or a different estimate the second time around. In the control condition, participants saw the description of the ultimatum game immediately followed by the outcome of the first trial, and then answered the questions shown above just once.

In this scenario invariance is not normatively required because the outcome of the first trial suggests that the responder is sensitive to the fairness of offers. Thus,  $Pr_2(A|O) = Pr_1(A|O,t) < Pr_1(A|O)$ , where  $t$  is the experience of knowing the outcome of the first trial.

## Results

**Average responses.** In each scenario we use  $Pr_1$  to refer to probability estimates before the experience and  $Pr_2$  for estimates after the experience. Average probabilities are shown in Table 2.

Table 2: Average estimates in Experiment 2 (standard deviations in parentheses).

lottery	$Pr(C)$	$Pr(W)$	$Pr(C W)$	$Pr(C \bar{W})$
$Pr_1$	0.29(0.22)	0.01(0.01)	0.77(0.22)	0.21(0.14)
$Pr_2$	0.33(0.25)	0.03(0.02)	0.76(0.20)	0.20(0.15)
Control	0.31(0.19)	0.04(0.04)	0.70(0.29)	0.29(0.20)
ultimatum	$Pr(A)$	$Pr(O)$	$Pr(A O)$	$Pr(A \bar{O})$
$Pr_1$	0.65(0.22)	0.61(0.29)	0.75(0.19)	0.44(0.30)
$Pr_2$	0.68(0.22)	0.74(0.24)	0.72(0.20)	0.38(0.30)
Control	0.62(0.25)	0.63(0.28)	0.70(0.25)	0.36(0.29)

**Control versus experimental conditions.** We found no reliable differences between  $Pr_2$  and the control estimates for each scenario using independent-sample  $t$ -tests. Thus, experimental participants seem not to have been influenced by having to evaluate the same probabilities twice.

**Analysis of the lottery scenario.** We first determined whether  $Pr(W)$  increased after participants saw the results of

the first four draws. As expected,  $Pr_2(W)$  was reliably larger than  $Pr_1(W)$  [paired  $t(49) = 7.7, p < .01$ , Wilcoxon  $p < .01$ ].

To see whether invariance holds, we compared  $Pr_1(C|W)$  to  $Pr_2(C|W)$  via paired  $t$ -test and found no reliable difference [ $df = 49, p > .05$ ]. Of the 50 participants, only 12 offered different values for  $Pr_2(C|W)$  versus  $Pr_1(C|W)$ . The 12 non-invariant participants made highly variable estimates, with average signed difference of 0.15 between  $Pr_1(C|W)$  and  $Pr_2(C|W)$  ( $SD = .57$ ), not reliably different from 0 [ $t(11) = 0.9, p > .05$ ].  $Pr_1(C|\bar{W})$  was likewise found to be close to  $Pr_2(C|\bar{W})$  [paired  $t(49) = 1.0, p > .05$ ]. Fifteen out of 50 participants gave a different  $Pr_2(C|\bar{W})$  estimate from  $Pr_1(C|\bar{W})$ . For these 15, the average signed difference between  $Pr_1(C|\bar{W})$  and  $Pr_2(C|\bar{W})$  was 0.04 ( $SD = 0.12$ ), again not reliably different from 0 [ $t(14) = 1.0, p > .05$ ]. The average invariance violation was only 1.42% with a median violation of 0. The mean invariance violation for  $Pr(C|\bar{W})$  was 1.33% with a median of 0. Since we obtained no estimate for  $Pr(W|C)$ , converse movement was not computed.

From the results above, invariance seems to hold rather well. For each participant, we therefore computed  $Pr_{total}(C)$  via (1) and  $Pr_{Jeff}(C)$  via (3), with  $G$  and  $B$  substituted by  $C$  and  $W$ . These estimates were compared to the participant's direct evaluation of  $Pr_2(C)$  via absolute difference. The means for total and Jeffrey error were .15 and .13, respectively, not reliably different via paired  $t$ -test [ $t(49) = 1.5, p > .05$ ] or Wilcoxon test ( $p > .05$ ).

**Analysis of the ultimatum game scenario.** We first determined whether  $Pr(O)$  increased after the reported rejection in the preceding trial. As expected,  $Pr_2(O)$  was reliably larger than  $Pr_1(O)$  [paired  $t(49) = 4.2, p < .01$ , Wilcoxon  $p < .01$ ].

Of the 50 participants, 33 offered a different  $Pr_2(A|O)$  estimate from  $Pr_1(A|O)$ . Thirty-eight out of 50 participants gave a different  $Pr_2(A|\bar{O})$  estimate from  $Pr_1(A|\bar{O})$ . The average invariance violation was 18.71% with a median of 12%. As expected, this violation was reliably greater than that in the lottery case (1.42%) [paired  $t(49) = 3.7, p < .01$ ]. For  $Pr(A|\bar{O})$  the mean invariance violation was 17.38%, reliably greater than  $Pr(C|\bar{W})$  in the lottery case (1.33%) [paired  $t(49) = 2.9, p < .01$ ]. Thus, invariance was violated to a greater extent here than in the lottery scenario.

## Discussion of Experiment 2

In the lottery scenario, invariance held for a majority of participants, and the absolute difference between  $Pr_2(C|W)$  and  $Pr_1(C|W)$  as a percentage of  $Pr_1(C|W)$  was quite small. When used to estimate  $Pr_2(C)$  via the law of total probability, we saw that  $Pr_1(C|W)$  could be substituted for  $Pr_2(C|W)$  with little loss of accuracy. In the ultimatum scenario, however, a majority of participants gave different estimates for  $Pr_2(A|O)$  compared to  $Pr_1(A|O)$  after learning the outcome of the first trial. Thus, invariance seems not to hold for the ultimatum scenario, as it ought not on normative grounds.<sup>1</sup>

<sup>1</sup>Without giving details, we note that Experiment 2 was repeated with 330 participants recruited over the internet via *Amazon Turk*.

## General Discussion

In Experiment 1, experience with the light changed the probability that the chosen card was blue, but had only mild impact on the probability of the giraffe *given that* the card was blue. That is,  $Pr_2(G|B) \approx Pr_1(G|B)$  as well as  $Pr_2(G|\bar{B}) \approx Pr_1(G|\bar{B})$ . These results conform to Jeffrey (1983)'s invariance requirement for updating a distribution on the basis of events whose probabilities are modified without reaching certainty. As a result, the updated probability  $Pr_2(G)$  was equally well predicted from the law of total probability on the basis of  $Pr_1(G|B)$  and  $Pr_1(G|\bar{B})$  versus  $Pr_2(G|B)$  and  $Pr_2(G|\bar{B})$ .

The invariance documented in Experiment 1 was selective inasmuch as greater movement was seen between the converse probabilities  $Pr_1(B|G)$  and  $Pr_2(B|G)$  than between  $Pr_1(G|B)$  and  $Pr_2(G|B)$ . The difference in movement makes normative sense because the giraffe is conditionally independent of the light given the color of the card whereas the color of the card is not conditionally independent of the light given the giraffe. Experiment 1 thus provides evidence that the participants were sensitive to the normative appeal of Jeffrey's rule, distinguishing (at least partially) between situations where it legitimately applies and where it does not.

The same conclusion is suggested by the results of Experiment 2. Only one of the two scenarios — involving the state lottery rather than the Ultimatum game — gave grounds for invariance, and participants honored the principle more in the lottery context. In the latter setting,  $Pr_2(C)$  (the revised probability of a car purchase) was predicted equally well from the law of total probability based on  $Pr_1(C|W)$  and  $Pr_1(C|\bar{W})$ , as it was from  $Pr_2(C|W)$  and  $Pr_2(C|\bar{W})$ .

Although the experiments support the hypothesis of (tacit) respect for Jeffrey's rule, the fact remains that a majority of our participants (51 of 100) changed their estimate of  $Pr(G|B)$  or  $Pr(C|W)$  between times 1 and 2. A slightly larger majority (54 of 100) did so for  $Pr(G|\bar{B})$  or  $Pr(C|\bar{W})$ . [The events *G*(iraffe), *B*(lue), *C*(ar), and *W*(in) come from the flashlight and lottery scenarios, where invariance is warranted.] In percentage terms, these shifts were sizeable in Experiment 1 (averaging around 47%) although much smaller in Experiment 2 (less than 2%). The psychology of updating is incomplete without an explanation of why invariance is not respected scrupulously in settings where it seems to be required normatively.

The mere fact of evaluating the same probabilities twice might explain some of the violation of invariance. The results of our control conditions, however, suggest that this effect was minor. Recall that control participants responded just once, in the phase 2 setting (e.g., after the light), yet produced estimates that were not reliably different from those gathered in phase 2 of the experimental condition. Another source of invariance violation might be illicit conversion of conditional probability statements, e.g., evaluating  $Pr(B|G)$  in place of the requested  $Pr(G|B)$ . This explanation is consistent with

---

The results were in close agreement with those reported here.

studies that highlight such conversion (as in Dawes, Mirels, Gold, and Donahue (1993)), but inconsistent with our own examination of conditional probability judgments (Zhao, Shah, & Osherson, 2009) in which little conversion was observed. Perhaps the variability of previous findings about conversion is somehow connected to the difference between the flashlight and lottery studies in their conformity to invariance.

A third possibility is that the flashlight procedure drew attention to the color dimension of the stimulus, reminding the participant of its predictive value. This realization might have been translated into more extreme conditional probabilities (higher for blue, lower for purple). The less vivid experience in the lottery scenario (merely being told about the first four numbers) would have had less impact, explaining the difference between the two experiments. As an alternative to vivacity, the greater impact of the flashlight might be related to the ineffable character of sensory impressions (as stressed by Jeffrey (1983, §11.1)); there is no such difficulty for the event of matching the first four lottery numbers. Of course, more data are needed to test hypotheses such as these.

## 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.

Doring, F. (1999). Why bayesian psychology is incomplete. *Philosophy of Science*, 66, 379–389.

Gardenfors, P. (1988). *Knowledge in flux: Modeling the dynamics of epistemic states*. Cambridge MA: MIT Press.

Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge UK: Cambridge University Press.

Harman, G. (1999). *Reasoning, meaning and mind*. Oxford UK: Oxford University Press.

Jeffrey, R. C. (1983). *The logic of decision* (2nd edition). Chicago IL: The University of Chicago Press.

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

Osherson, D. (2002). *Order dependence and jeffrey conditionalization*. (Available via <http://www.princeton.edu/osherson/papers/jeff3.pdf>)

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo, CA: Morgan Kaufmann.

Stanovich, K. (2010). *Decision making and rationality in the modern world*. Oxford UK: Oxford University Press.

Wagner, C. G. (2002). Probability kinematics and commutativity. *Philosophy of Science*, 69, 266–278.

Williams, P. M. (1980). Bayesian conditionalisation and the principle of minimum information. *British Journal for the Philosophy of Science*, 31, 131–144.

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