

# How much ambiguity aversion?

## Finding indifferences between Ellsberg's risky and ambiguous bets

Ken Binmore · Lisa Stewart · Alex Voorhoeve

Published online: 28 November 2012  
© Springer Science+Business Media New York 2012

**Abstract** Experimental results on the Ellsberg paradox typically reveal behavior that is commonly interpreted as ambiguity aversion. The experiments reported in the current paper find the objective probabilities for drawing a red ball that make subjects indifferent between various risky and uncertain Ellsberg bets. They allow us to examine the predictive power of alternative principles of choice under uncertainty, including the objective maximin and Hurwicz criteria, the sure-thing principle, and the principle of insufficient reason. Contrary to our expectations, the principle

---

Parts of the experimental data were gathered while Lisa Stewart was a researcher in the Harvard Psychology Department and Alex Voorhoeve was a Faculty Fellow at Harvard's Safra Center for Ethics. We thank the Decision Science Laboratory at Harvard and the ELSE laboratory at University College London for the use of their facilities, and the Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD) for financial support. Ken Binmore thanks the British Economic and Social Research Council through the Centre for Economic Learning and Social Evolution (ELSE), the British Arts and Humanities Research Council through grant AH/F017502 and the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant 295449. Alex Voorhoeve thanks the Safra Center for Ethics for its Faculty Fellowship and the British Arts and Humanities Research Council through grant AH/J006033/1. Results were presented at Bristol University, the European University Institute in Florence, Harvard University, the LSE, the University of Siena, and the University of York (UK). We thank Richard Bradley, Barbara Fasolo, Joshua Greene, Glenn Harrison, Jimmy Martinez, Katie Steele, Joe Swierzbinski, Peter Wakker, the Editors of and an anonymous referee for the JRU, and those present at our seminars for their comments.

K. Binmore  
Economics Department, University College London, London, WC1E 6BT, UK

L. Stewart  
Psychology Department, Australian National University, Canberra, ACT 0200, Australia

A. Voorhoeve (✉)  
Department of Philosophy, Logic and Scientific Method, London School of Economics,  
London, WC2E 2AE, UK  
e-mail: a.e.voorhoeve@lse.ac.uk

of insufficient reason performed substantially better than rival theories in our experiment, with ambiguity aversion appearing only as a secondary phenomenon.

**Keywords** Ambiguity aversion · Ellsberg paradox · Hurwicz criterion · Maximin criterion · Principle of insufficient reason

**JEL Classification** C91 · D03

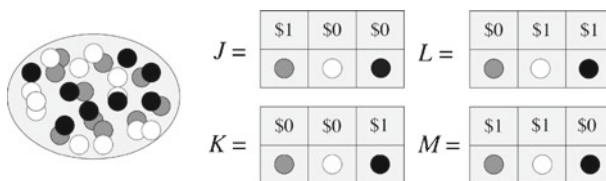
## 1 Ellsberg paradox

Ellsberg (1961) famously proposed an experiment the results of which have become known as the “Ellsberg paradox” because they are inconsistent with the predictions of expected utility theory.

In one version of Ellsberg’s experiment, a ball is drawn from an urn containing ten red balls and twenty other balls, of which it is known only that each is either black or white. The gambles  $J$ ,  $K$ ,  $L$ , and  $M$  in Fig. 1 represent various reward schedules depending on the color of the ball that is drawn. Ellsberg predicted that most people will strictly prefer  $J$  to  $K$  and  $L$  to  $M$ . However, if the probabilities of picking a red, black, or white ball are respectively  $R$ ,  $B$ , and  $W$ , then the first preference implies that  $R > B$  and the second that  $B > R$ . So such behavior cannot be consistent with maximizing expected utility relative to any subjective probability distribution.

Ellsberg’s explanation for such violations of Bayesian decision theory is that the subjects’ choices show an aversion to ambiguity for which the theory makes no provision. The subjects know there is a probability that a black ball will be chosen, but this probability might be anything between 0 and  $\frac{2}{3}$ . When they choose  $J$  over  $K$ , they reveal a preference for winning with a probability that is certain to be  $R = \frac{1}{3}$  rather than winning with a probability  $B$  that might be anything in the range  $[0, \frac{2}{3}]$ . When they choose  $L$  over  $M$ , they reveal a preference for winning with a certain probability of  $1 - R = W + B = \frac{2}{3}$  to winning with a probability  $1 - W = R + B$  that might be anything in the range  $[\frac{1}{3}, 1]$ .

*Previous experimental evidence* For the version of the experiment outlined above, Ellsberg’s prediction that a majority of subjects would display ambiguity aversion



**Fig. 1** Ellsberg paradox: in the version illustrated an urn contains ten red balls and another twenty balls of which it is only known that they are either black or white. A ball is chosen at random from the urn, the color of which determines the award of a prize (which is here taken to be one dollar). Whether subjects win or lose depends on which of the lottery tickets  $J$ ,  $K$ ,  $L$ , or  $M$  they are holding

has generally been confirmed.<sup>1</sup> However, when one varies the original set-up, the picture is more complex. In the space of gains, ambiguity aversion is prevalent. However, in the space of losses, some studies find that ambiguity-loving behavior dominates (Wakker 2010, p. 354). The literature also finds that the extent and strength of ambiguity aversion depends on whether the known probability of winning is high or low (with ambiguity aversion being stronger for moderate to high probabilities) and whether the ambiguous option is presented alone or juxtaposed to a risky option (with the effect being stronger in the latter case).<sup>2</sup> In sum, a great deal remains to be learned about the impact of framing effects on human decision behavior in uncertain situations.

*New experiment* Our experiment varies the classic design by changing the numbers of balls so that the probability that a red ball is chosen is altered from  $\frac{1}{3}$  to  $R$ . We then estimate the value  $r_1$  of  $R$  that makes a particular subject indifferent between  $J$  and  $K$ . We also estimate the value  $r_2$  of  $R$  that makes the same subject indifferent between  $L$  and  $M$ .<sup>3</sup>

A subject who honors Savage's (1954, p. 21) sure-thing principle will have

$$r_1 = r_2. \quad (1)$$

The rationale in our special case is that, in comparing  $J$  and  $K$  and in comparing  $L$  and  $M$ , what happens if a white ball is drawn is irrelevant. The two comparisons therefore depend only on what happens if a white ball is not drawn. But if a red or black ball is sure to be drawn, then  $J$  is the same as  $M$  and  $K$  is the same as  $L$ . So  $J$  and  $K$  should be regarded as being worth the same if and only if the same is true of  $L$  and  $M$ .

Given the widespread finding of ambiguity aversion in the classic Ellsberg paradox, our aim in this paper was to examine the extent to which the Hurwicz criterion discussed in the next section explains deviations from the sure-thing principle and other tenets of Bayesian decision theory caused by ambiguity aversion.

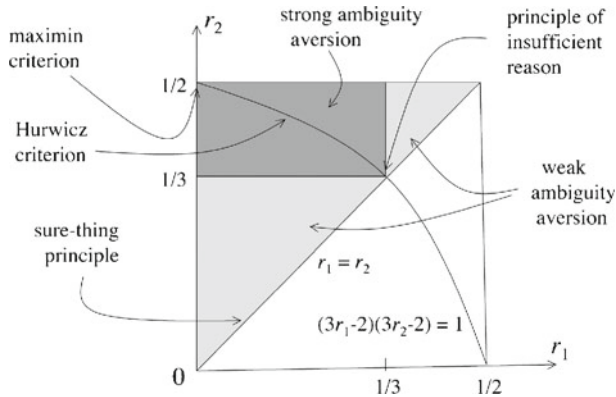
## 2 Theories

We consider various theories of choice behavior under uncertainty. Figure 2 illustrates the behavior predicted by each of the principles reviewed. These all have

<sup>1</sup>See the overview of the evidence presented by Camerer and Weber (1992). See also Fox and Tversky (1995), Keren and Gershten (1999), and Liu and Colman (2009).

<sup>2</sup>Ahn et al. (2010), Chow and Sarin (2001), Curley and Yates (1989), Hogarth and Einhorn (1990), Etner et al. (2012), Fox and Tversky (1995), Halevy (2007), Hey et al. (2010), Hsu et al. (2005), Trautmann et al. (2008), Wakker (2010), and Abdellaoui et al. (2011). The link <http://aversion-to-ambiguity.behaviouralfinance.net/> leads to many more papers on ambiguity aversion published since the year 2000.

<sup>3</sup>In manipulating the number of red balls to determine the extent of ambiguity aversion, our approach resembles that of MacCrimmon and Larsson (1979), Kahn and Sarin (1988), Viscusi and Magat (1992), and Viscusi (1997), each of which finds substantial ambiguity aversion. Our approach differs from the latter three studies in that rather than asking individuals outright how much the known probability would need to be to render them indifferent between a risky and uncertain bet, we use their choices between bets to estimate an indifference interval.



**Fig. 2** Theoretical possibilities: Bayesian decision theory predicts an outcome on the diagonal of the square. The principle of insufficient reason says that the probabilities of a black or a white ball should be taken to be equal, as there is no reason to suppose one more likely than the other. Bayesian decision theory then predicts that  $r_1 = r_2 = \frac{1}{3}$ . An observation with  $r_2 > r_1$  in the lightly shaded region is taken to be a case of weak ambiguity aversion. An observation with  $r_1 < \frac{1}{3}$  and  $r_2 > \frac{1}{3}$  in the more deeply shaded region is taken to be a case of strong ambiguity aversion

variants that apply when the outcomes are not merely winning or losing as in our paper. For example, the Hurwicz criterion has been generalized to what is sometimes called alpha-max-min (Ghirardato and Marinacci 2002). Our case is simpler because it leaves no room to maneuver about the nature of the utility function that can be attributed to a subject. We need only consider the (Von Neumann and Morgenstern) utility function that assigns a value of 0 to losing and 1 to winning. Difficulties about the level of a subject's risk aversion and the like therefore do not arise in our experiment.

*Maximin criterion (MXN)* The most widely discussed alternative to Bayesian decision theory was proposed by Wald (1950) and has been developed since that time by numerous authors (see Gilboa 2004). It sometimes goes by the name of the maximin criterion, because it predicts that subjects will proceed as though they are facing the least favorable of all the probability distributions that ambiguity allows.

We consider Von Neumann's version of the maximin criterion here.<sup>4</sup> In this objective version, the decision-maker's ambiguity is assumed to extend to all probability distributions that are consistent with the objective data.

<sup>4</sup>The objective maximin criterion is often referred to as the minimax criterion. The confusion between maximin and minimax presumably arises because minimax equals maximin in Von Neumann's famous minimax theorem. The confusion is sometimes compounded because Savage (1954) proposed a further decision criterion called the *minimax regret criterion*, which happens to make the same predictions as the maximin criterion in the special case considered in this paper. (Savage (1954, p. 16) distinguished between large and small worlds, recommending his minimax regret criterion for the former. He variously describes using Bayesian decision theory outside a small world as "preposterous" and "utterly ridiculous".)

Applying the objective maximin criterion to our example, we are led to seek to maximize a utility function defined by  $u(J) = R$ ,  $u(K) = 0$ ,  $u(L) = 1 - R$ , and  $u(M) = R$ . So for  $0 < R < \frac{1}{2}$ ,  $J$  is preferred to  $K$  and  $L$  to  $M$ .

The literature also features a subjective version of the maximin criterion in which the decision-maker may have internal reasons for excluding some of the probability distributions taken into account by the objective version (Gilboa and Schmeidler 1989). Applying the subjective maximin criterion in the extreme case when the decision-maker considers only a single probability distribution gives the same predictions as the sure-thing principle. If we insist on some ambiguity, the class of subjective versions of the maximin criterion coincides with the notion of weak ambiguity aversion to be considered later.

*Hurwicz criterion (HWZ)* The Hurwicz criterion (1951) goes back to the beginnings of decision theory. Hurwicz’s proof was simplified by Arrow and Hurwicz (1972). As a consequence, it is sometimes referred to as the Arrow–Hurwicz criterion. It features in Luce and Raiffa’s (1957) discussion of decisions made under complete ignorance in their book *Games and Decisions*. Ellsberg (2001) considers it at length in his *Risk, Ambiguity, and Decision*, both as a normative criterion and as an explanation for ambiguity-averse choices. Binmore (2009, p. 166) offers a normative defense of a multiplicative version of the Hurwicz criterion, which would be indistinguishable from the standard (additive) Hurwicz criterion in our experiment. For a review of the criterion’s role in the literature, see Etner et al. (2012).

At an early stage, Milnor (1954) offered axioms that characterize several of the decision criteria considered in this paper. In the case of the Hurwicz criterion, he replaces a version of the standard independence axiom, which he calls *column linearity*, by a new axiom that he calls *column duplication*. This axiom says that the decision maker’s choice will remain unchanged if a new state is appended from which the same consequences as an existing state would follow for every choice available to the decision maker. The idea is that a completely ignorant decision maker will have no reason not to assimilate the new state into the existing state that it duplicates.

We follow Milnor in regarding the Hurwicz criterion as applicable in situations of partial ignorance only after the decision maker’s partial knowledge has been incorporated into her model in terms of objective upper and lower probabilities that bound the possible range of the probability  $R$  that a red ball is drawn. Subjective versions of the Hurwicz criterion are possible, but we only consider the objective version.

The standard Hurwicz criterion balances the pessimism of the objective maximin criterion against the optimism of what might be called the objective maximax criterion. The Hurwicz criterion values a gamble  $G$  offering a prize with probability  $P$  with the utility function

$$u(G) = (1 - h)\underline{P} + h\overline{P}, \tag{2}$$

where  $[\underline{P}, \overline{P}]$  is the (objective) range of possible values of  $P$ , and the constant  $h$  ( $0 \leq h \leq 1$ ) registers how averse the subject is to ambiguity. The case  $h = 0$  of maximal aversion coincides with the objective maximin criterion. The case  $h = 1$  corresponds to the objective maximax criterion. The case  $h = \frac{1}{2}$  is indistinguishable from the principle of insufficient reason in our experiment.

In our example, the Hurwicz criterion yields  $u(J) = R$ ,  $u(K) = h(1 - R)$ ,  $u(L) = 1 - R$ , and  $u(M) = (1 - h)R + h$ . If a subject is indifferent between  $J$  and  $K$  when  $R = r_1$ , it follows that  $h = r_1/(1 - r_1)$ . Similarly, if the subject is indifferent between  $L$  and  $M$  when  $R = r_2$ , then  $h = (1 - 2r_2)/(1 - r_2)$ . Assuming that the same value of  $h$  applies in both cases, it follows that the Hurwicz criterion predicts that  $r_1$  and  $r_2$  are connected by the equation<sup>5</sup>

$$(3r_1 - 2)(3r_2 - 2) = 1. \quad (3)$$

*Sure-Thing Principle (STP)* The Hurwicz criterion needs to be compared with the orthodox Bayesian approach (expected utility theory), which denies that subjects might be unable to resolve ambiguities about what subjective probability to assign to events. In this case,  $u(J) = R$ ,  $u(K) = B$ ,  $u(L) = W + B$ , and  $u(M) = R + W$ . So the criteria for indifference between  $J$  and  $K$  and between  $L$  and  $M$  are the same:  $r_1 = R = B = r_2$ .

We have already seen that we need no more than the sure-thing principle to justify the conclusion that  $r_1 = r_2$ , which one might also categorize as representing ambiguity neutrality.

*Laplace's Principle of Insufficient Reason (PIR)* This principle says that events should be assigned the same subjective probability if no reason can be given for regarding one event as more likely than another. A subject who believes this to be true of drawing a black or a white ball will therefore assign them equal probability, so that  $W = B$ . When this result is combined with the equation  $r_1 = R = B = r_2$ , we obtain that

$$r_1 = r_2 = \frac{1}{3}. \quad (4)$$

This outcome is also predicted by the Hurwicz criterion with  $h = \frac{1}{2}$ .

*Weak Ambiguity Aversion (WAA)* We say that  $r_1 < r_2$  indicates weak ambiguity aversion, because it implies that  $J \succ K$  and  $L \succ M$  when the probability  $R$  of a red ball being drawn lies between  $r_1$  and  $r_2$ . Reversing the inequality generates a criterion for weak ambiguity-loving behavior.

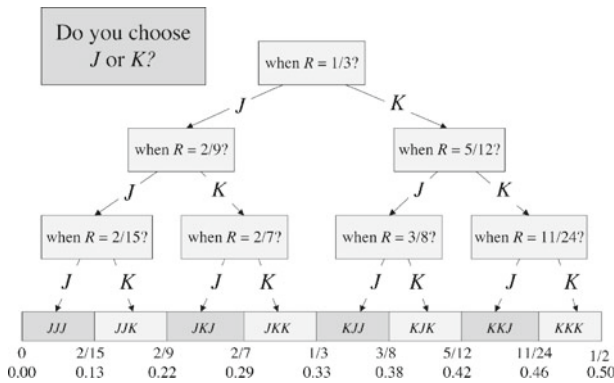
Outcomes that satisfy the Hurwicz criterion with  $h < \frac{1}{2}$  are ambiguity averse in both the weak sense and the strong sense that follows.

*Strong Ambiguity Aversion (SAA)* We treat pairs  $(r_1, r_2)$  with  $r_1 < \frac{1}{3}$  and  $r_2 > \frac{1}{3}$  as cases of strong ambiguity aversion. To see why, recall that a Bayesian subject will express indifference between  $J$  and  $K$  when  $R = B$ . So if  $W = B$ , then  $r_1 = \frac{1}{3}$ . If one regards behaving as though  $B < W$  as a manifestation of strong ambiguity

<sup>5</sup>Binmore's (2009, p. 166) multiplicative version of the Hurwicz criterion also yields Eq. 3 to a first order of approximation. To a second order of approximation, it yields

$$(3r_1 - 2)(3r_2 - 2) = 1 + c(r_1 - r_2)\{2(1 - r_1)(1 - r_2) - 1\},$$

for some small positive constant  $c$ .



**Fig. 3** A titration: the tree shows the questions asked about what happens when the probability of RED is  $R$  in order to locate  $r_1$  in one of eight subintervals of  $[0, \frac{1}{2}]$ . For example, someone who answers  $KJJ$  is assigned a value of  $r_1$  satisfying  $\frac{1}{3} \leq r_1 \leq \frac{3}{8}$ . The points dividing the subintervals are chosen to be rational numbers so that the questions can be framed in terms of decks of cards. The pairs  $(0, \frac{1}{2})$ ,  $(\frac{2}{15}, \frac{11}{24})$ ,  $(\frac{2}{9}, \frac{5}{12})$ ,  $(\frac{2}{7}, \frac{3}{8})$ , and  $(\frac{1}{3}, \frac{1}{3})$  lie on the curve  $(3r_1 - 2)(3r_2 - 2) = 1$ . The same tree is used to locate  $r_2$ , except that  $L$  replaces  $K$  and  $M$  replaces  $J$

aversion, then  $r_1 < \frac{1}{3}$ . The requirement that  $r_2 > \frac{1}{3}$  is derived similarly. Reversing all inequalities generates a criterion for strong ambiguity-loving behavior.

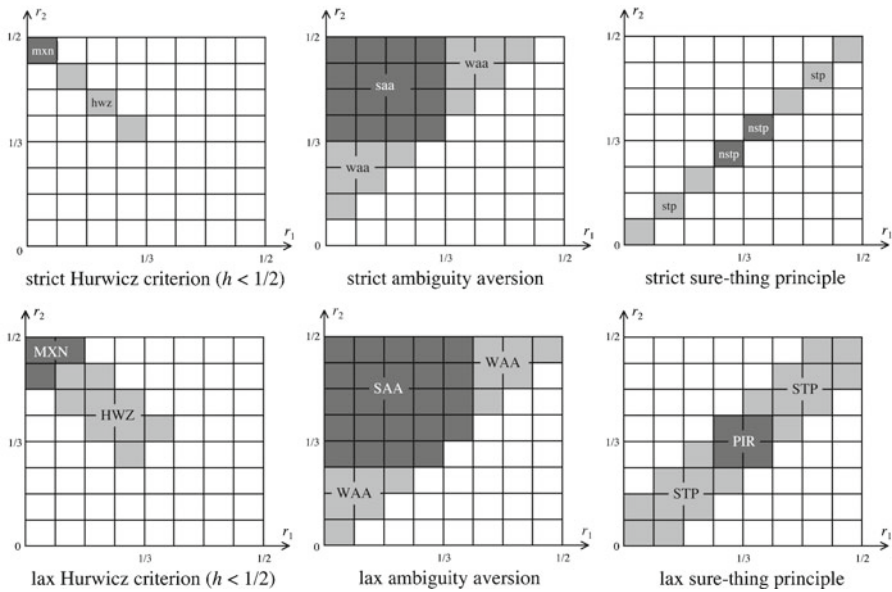
### 3 Experiments

We followed the practice common in psychology of seeking to estimate the indifference probabilities  $r_1$  and  $r_2$  using a titration.<sup>6</sup> Subjects were asked a sequence of questions about their choices between  $J$  and  $K$ , and between  $L$  and  $M$  for various values of the probability  $R$  that a red ball will be drawn. The aim of this titration is to locate the values of  $r_1$  and  $r_2$  within eight subintervals of  $[0, \frac{1}{2}]$  using the scheme illustrated in Fig. 3.<sup>7</sup>

Our titration locates an estimate of a subject’s  $(r_1, r_2)$  within one of 64 squares of a chessboard. Figure 4 shows the regions on this chessboard that we shall regard as providing support for the various decision theories we consider. The regions identified in the top row do not allow for subject error. The regions identified in the bottom

<sup>6</sup>Psychologists favor the use of a titration over simply asking subjects for their indifference probabilities, but there is a risk that subjects might not always answer the titration questions truthfully because they prefer being paid on Ellsberg bets that can only be reached by lying. The monetary payoffs associated with each bet were chosen to make such misrepresentation unprofitable for risk-neutral subjects who honor the principle of insufficient reason, but ambiguity-averse subjects could sometimes gain by lying. However, it would be necessary for them first to learn what future questions would be asked before they could exploit the opportunity for misrepresentation, and we found no significant evidence of learning in comparing subjects’ behavior in later stages of the experiment.

<sup>7</sup>In this exercise, it is assumed that  $K > J$  when  $R = 0$ , and  $J \geq K$  when  $R \geq \frac{1}{2}$ ; also  $L > M$  when  $R = 0$ , and  $M \geq L$  when  $R \geq \frac{1}{2}$ .



**Fig. 4** Regions of interest: the results are reported using chessboards showing the percentage of times that  $(r_1, r_2)$  is observed in one of 64 possible squares. The chessboards in the *top row* of the figure correspond to a strict interpretation of each region, which allows no margin for mistaken choices by the subjects. (The regions of interest differ somewhat from Fig. 2 because the constituent squares of the table are not drawn to scale.) The chessboards in the *bottom row* allow a margin for subject error. For each chessboard in the left column, the whole shaded region corresponds to the Hurwicz criterion (hwz or HWZ) with an ambiguity-averse coefficient. The more deeply shaded region corresponds to the maximin criterion (mxn or MXN). In the *middle column*, the whole shaded region corresponds to weak ambiguity aversion (waa or WAA). The more deeply shaded region corresponds to strong ambiguity aversion (saa or SAA). In the *right column*, the whole shaded region corresponds to the sure-thing principle (stp or STP). The more deeply shaded region in the lax chessboard corresponds to the principle of insufficient reason (PIR). There is no corresponding exact region pir because this would be identical to PIR. Instead, we distinguish the (neutral) part of stp that lies in PIR as nstp

row permit a margin for error that amounts to allowing a subject at most one careless choice that results in either  $r_1$  or  $r_2$  (but not both) being placed in an interval adjacent to the interval in which it would have been placed if the choice had been made carefully.<sup>8</sup>

The principle of insufficient reason (PIR) requires special treatment because  $\frac{1}{3}$  is an endpoint of two of our intervals. Any value of  $(r_1, r_2)$  that lies in one of the four squares of our chessboard with a corner at  $(\frac{1}{3}, \frac{1}{3})$  is therefore treated as supporting PIR. This region is not expanded to include subjects' choices in intervals adjacent to these four squares because subjects who intended to act in conformity with PIR

<sup>8</sup>For example, a subject who reveals values for  $r_1$  and  $r_2$  that both lie in the interval  $[\frac{1}{3}, \frac{3}{8}]$  (corresponding to the choices *KJJ* and *LMM*) will be regarded as satisfying the sure-thing principle (stp or STP). But our lax criterion also includes in STP a subject whose value of  $r_1$  is the same, but whose value of  $r_2$  lies in the neighboring interval  $[\frac{2}{7}, \frac{1}{3}]$  corresponding to the choice *MLL* that might result if the subject made a misjudgment at the first question but answered later questions accurately.

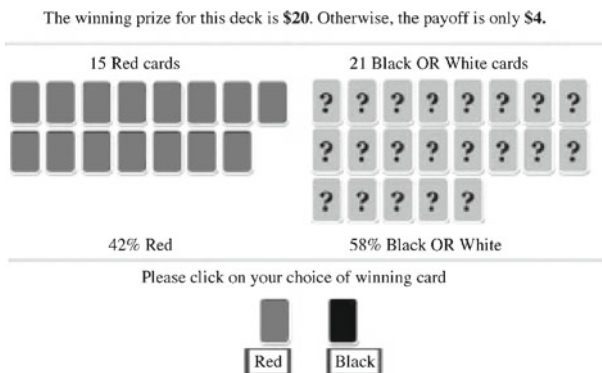


would be indifferent between  $J$  and  $K$  (and between  $L$  and  $M$ ) for  $r_1 = \frac{1}{3}$ . It follows that if their choices were to place them in a square adjacent to the four squares that have  $[\frac{1}{3}, \frac{1}{3}]$  as a midpoint, then they would have strayed further from their true preference than subjects who intended to act in accordance with one of the other principles and who ended up in a square adjacent to a region permitted by that principle. For similar reasons, we do not shrink PIR to obtain a smaller region  $\text{pir}$ . The region  $\text{nstp}$  in Fig. 4 should be thought of only as the neutral part of  $\text{stp}$ .

It will be necessary to consider the extent to which apparent support for one theory needs to be assessed in the light of the support which the same data gives an alternative theory. For example, a fraction of the data that is consistent with weak ambiguity aversion (WAA) also supports the principle of insufficient reason (PIR). In considering such issues, we use the notation  $\text{WAA} \setminus \text{PIR}$ , which consists of all squares on the chessboard in the region WAA but not in the region PIR.

*Shuffling and dealing* In our experiments, urns of colored balls were replaced by decks of colored cards. Seated in front of a computer screen, subjects chose whether to bet on  $J$  or  $K$  (or whether to bet on  $L$  or  $M$ ) for various values of the probability  $R$  of drawing a red card. For example, Fig. 5 shows the screen the subjects saw when being asked whether they want to bet on  $J$  or  $K$  when  $R = \frac{5}{12}$ . For full details of the experimental interface (and all the results) see the link (note that the preferred browsers are Firefox [Version 12.0 or later] and Safari): <http://alkami.org/ells/>.

*Allaying suspicion* An on-screen introduction explained the structure of the experiment and the nature of the choices subjects would face. Subjects were told that they would choose between bets like  $J$  or  $K$  in ignorance of the mixture of BLACK and WHITE cards. Special care was taken to illustrate the nature of this ignorance. Subjects were given an illustrative deck of 6 RED cards and 15 BLACK OR WHITE cards, the latter marked on the screen with a “?” on the back. They were then told that the “?” cards could be any mixture of BLACK and WHITE cards, with three subsequent



**Fig. 5** Experimental interface: when confronted with this particular interface, the subject is being asked whether he or she prefers  $J$  or  $K$  when the probability of drawing a red card is  $R = \frac{5}{12}$

screens inviting them to move the mouse over the “?” cards, revealing three illustrative mixtures, under the headings: “It could be that all cards that are NOT RED are BLACK”; “It could be that all cards that are NOT RED are WHITE”; and “It could be that all cards that are NOT RED are any of the many possible mixtures of BLACK and WHITE, for example ...”, with the example consisting of 5 WHITE and 10 BLACK cards.

Inspired by Hey et al. (2010), we sought to allay any suspicion of deceit on our part by making transparent the preparation of the decks from which a winning card would be drawn. After making two practice choices, subjects were invited to the front of the room to witness one of the practice bets being played for illustrative purposes only. The experimenter opened a box of RED cards and box of BLACK OR WHITE cards, and counted out the number of RED and BLACK OR WHITE cards in the first practice choice (respectively 6 and 15). These were placed in a card-shuffling machine to randomize the order of the deck. Finally, a subject exposed the third card from the top in the shuffled deck, the color of which determined whether subjects had won or lost. Subjects were told, truthfully, that of the subsequent 24 choices they faced in the real experiment, two bets would be randomly selected by the computer to be played for money in this manner at the end of the experiment, with the choices they had made determining the winning color(s). (This randomization was done for each subject individually, so that all subjects had tailor-made bets constructed and played for them.)

This procedure may be relevant to the relatively low levels of ambiguity aversion we observed. For example, Pulford (2009) reports significantly more ambiguity aversion after drawing attention to the possibility of experimental deceit. If a subject believes that the experimenter is deceitfully manipulating the shuffling-and-dealing protocols to minimize payoffs (or for some other reason), then the problem faced by the subject ceases to be a one-person decision problem and becomes instead a game played between the subject and the experimenter (Brewer 1965; Schneeweiss 1968; Kadane 1992). In an extreme case, the subject may (unconsciously?) perceive this game as zero-sum, in which rational play (according to Bayesian decision theory) requires the play of the subject’s maximin strategy (Schneeweiss 1968). Researchers are then at risk of misinterpreting such play as the use of an ambiguity-averse strategy in a one-person decision problem.<sup>9</sup>

*Subjects* Two types of subject were studied: on-site subjects and on-line subjects. The on-site participants were recruited from lists of volunteers maintained by the laboratories at which various versions of the experiment were run. These subjects were paid according to their success in selecting winning cards. On-line subjects participated from remote sites with negligible payment (and without the opportunity

---

<sup>9</sup>Savage (1954, p. 16) would perhaps have commented that leaving room for suspicion of dishonest manipulation by the experimenter creates a large world for the subjects. Our design is intended to make the subjects’ world small in this respect.

to check up on how the cards were shuffled and dealt).<sup>10</sup> The behavior of on-line subjects turned out to resemble that of on-site subjects, but is much noisier. (The hypothesis that on-line subjects paid less attention to their choice problem than on-site subjects is supported by the fact that the mean response time of on-line subjects to each choice was just over half the mean response time of on-site subjects.) The discussion that follows therefore focuses on the on-site data.

### 3.1 Version 1

Following the practice choices and demonstration session, the subjects returned to their screens and were taken through the experiment with the aid of a computerized interface. The on-site edition of Version 1 of the experiment was run at the Harvard Decision Science Laboratory, using the Harvard Psychology Department subject pool. A pilot that led to some minor design changes is not reported. The aim was to elicit from each subject four pairs of observations for the indifference intervals in which to locate  $r_1$  and  $r_2$ . The experiment had five stages:

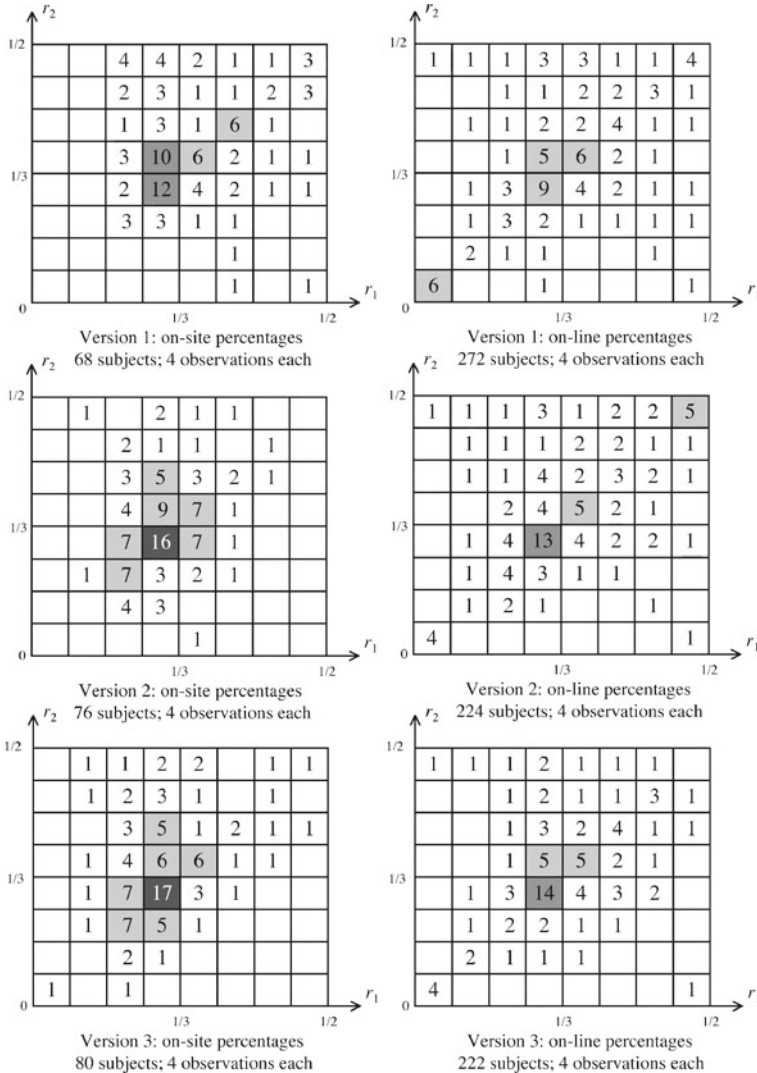
1. At the beginning of Round 1, subjects were told that the round had the following four parts, with the choices in each part being constructed from the same decks. Subjects were not told that the probability  $R$  of a red card would vary according to the titration of Fig. 3.
  - (a) Three choices between  $J$  (RED wins) and  $K$  (BLACK wins).
  - (b) Three choices between  $J$  and  $K'$  (WHITE wins).
  - (c) Three choices between  $L$  (WHITE and BLACK win) and  $M$  (RED and WHITE win). (Note that  $L$  was described as “NOT-RED wins”, and  $M$  as “NOT-BLACK wins”.)
  - (d) Three choices between  $L$  and  $M'$  (RED and BLACK win). (Note that  $L$  was described as “NOT-RED wins”, and  $M'$  as “NOT-WHITE wins”.)

Items (a) and (c) above determined one estimate of  $(r_1, r_2)$  to be compared with Fig. 4. Items (b) and (d) determined a second estimate.
2. Round 2 was identical to Round 1, save that YELLOW replaced WHITE and BLUE replaced BLACK. This round therefore provided a further two estimates of  $(r_1, r_2)$ .
3. Subjects were taken one-by-one through the questions from a Life Orientation Test as revised by Scheier et al. (1994).
4. Subjects were asked five questions about whether the tasks and questions had been clear, and whether any problems arose during the experiment.
5. Two gambles (one from each round) were chosen at random by the computer for each subject to be played for real. (The prizes were chosen to render subjects choosing according to the principle of insufficient reason indifferent about which

<sup>10</sup>We used Amazon’s Mechanical Turk (<https://www.mturk.com>), Psychological Research on the Net (<http://psych.hanover.edu/Research/exponnet.html>), and the Research Subject Volunteer Program (<http://alkami.org> <http://alkami.org>). Participants recruited via Mechanical Turk were paid a token \$0.05 for completing the approximately six-minute study.

of their choices were played for real.) On-site subjects left the laboratory with an average of \$22.

The data from Version 1 are displayed in Fig. 6 and summarized in Table 1. A detailed analysis follows below. Here, we note two things which surprised us. First, though there was some evidence for ambiguity aversion, there was not as much as our reading of the literature had led us to expect. Second, we were disappointed not to find much support for the Hurwicz criterion for any value of  $h$  except those around  $\frac{1}{2}$ , for which the principle of insufficient reason offers a competing explanation.



**Fig. 6** Summary of aggregate results. The on-site and on-line data for all observations for each version of the experiment are shown. Shaded squares indicate a particularly high concentration of responses

**Table 1** Summary of aggregate percentages

	0	aggregate percentages for on-site versions			aggregate percentages for on-line versions				0	aggregate percentages for on-site versions			aggregate percentages for on-line versions		
		1	2	3	1	2	3			1	2	3	1	2	3
POP	100	100	100	100	100	100	100	pop	100	100	100	100	100	100	100
MXN	5	1	1	1	2	2	2	mxn	2	0	0	0	1	1	1
HWZ ( $h < 1/2$ )	19	38	48	45	28	32	33	hwz ( $h < 1/2$ )	6	13	13	9	8	6	7
SAA	38	56	63	64	45	46	47	saa	25	32	27	28	20	20	19
WAA	56	74	75	80	72	73	73	waa	44	42	42	43	36	36	33
STP	34	60	69	65	62	64	66	stp	13	32	33	37	36	37	40
PIR	6	31	39	32	24	25	28	nspt	3	18	23	23	15	18	19
HWZ \ STP	14	11	15	16	8	10	9	stp \ nspt	9	15	10	14	21	19	22
SAA \ STP	30	25	20	27	20	20	18	strict criteria							
WAA \ STP	33	27	22	27	22	22	19	<div style="display: flex; justify-content: space-around;"> <div style="width: 15px; height: 15px; background-color: #cccccc; border: 1px solid black;"></div> <span>significant deviation at the 5% level from the null hypothesis toward the criterion</span> </div> <div style="display: flex; justify-content: space-around; margin-top: 5px;"> <div style="width: 15px; height: 15px; background-color: #333333; border: 1px solid black;"></div> <span>significant deviation at the 1% level from the null hypothesis toward the criterion</span> </div>							
STP \ PIR	28	29	29	33	38	39	38								
POP \ STP	100	100	100	100	100	100	100								
MXN \ STP	7	2	4	4	6	4	6								
HWZ \ STP	29	27	49	45	22	28	27								
SAA \ STP	45	64	65	77	54	54	52								
WAA \ STP	50	68	69	78	58	61	56								

lax criteria

The acronyms for different theories appear in Section 2. The upper part of the table for lax criteria shows percentages for each region of Fig. 4 of the total data (POP). The first column (0) shows what the percentages would be if all choices were made at random and the population were very large. The lower part of the table for lax criteria shows percentages of the data exclusive of data that falls in STP (POP \ STP). The shaded squares indicate percentages that are statistically significant at the 1% and 5% levels. (The means of the four observations from each subject are treated as independent for an application of the Normal approximation to the Binomial distribution)

### 3.2 Version 2

One possible explanation we considered was that the framing of the choices between *L* and *M* as between “NOT-BLACK WINS” and “NOT-RED WINS” affected the subjects. (The analogous ‘not’-description was given for the choices between *L* and *M'*.) In order to establish whether this was a key factor, we ran Version 2, which is exactly the same as Version 1 except for a change in how the problems faced by the subjects are framed, substituting the description “BLACK AND WHITE” for “NOT-RED”, “RED AND WHITE” for “NOT-BLACK”, and so on. The on-site version of this new experiment was carried out in the ELSE laboratory at University College London with 76 on-site subjects.<sup>11</sup>

Though, as our analysis in Section 5.2 shows, the results of the on-site subjects for Version 2 differed significantly in some respects from the results of the on-site

<sup>11</sup>The prizes were approximately equal to their previous dollar values but were denominated in British pounds. The average payout was around £13.

subjects for Version 1, they were similar in that they displayed only limited evidence of ambiguity aversion and provided little support for the Hurwicz criterion other than with  $h$  in the neighborhood of  $\frac{1}{2}$ .

### 3.3 Version 3

We remained surprised not to see more ambiguity aversion in Version 2. A possible explanation proposed by Raiffa (1961) is that on-site subjects could theoretically convert the problem into one in which an unambiguous distribution is available that assigns BLACK and WHITE equal probabilities if they believed our (true) claims that:

1. They would face the same decision tree for the RED versus BLACK, RED versus WHITE, and all subsequent versions of the  $J$  versus  $K$  and  $J$  versus  $K'$  choices (and the iterations of the  $L$  versus  $M$  and  $L$  versus  $M'$  choices);
2. The decks involved in these choices would all be constructed from the same set of BLACK OR WHITE (or BLUE OR YELLOW) decks;
3. Each of the subjects' choices was equally likely to be played for real.

No appeal to the principle of insufficient reason is then necessary to justify playing according to its tenets. To see why, consider the strategy of choosing BLACK when offered the choice between RED and BLACK, and choosing WHITE when offered the analogous choice between RED and WHITE. For a subject who held the aforementioned three beliefs, in Versions 1 and 2 of our experiment, this strategy is equivalent to turning down RED in favor of an equiprobable lottery between BLACK and WHITE, with a probability  $\frac{1}{3}$  of winning.

Consider next the strategy of choosing BLACK when offered the choice between RED and BLACK, and choosing RED AND WHITE when offered the choice between RED AND WHITE and BLACK AND WHITE. For a subject who held the aforementioned three beliefs, this is equivalent to turning down an equiprobable lottery between RED and BLACK AND WHITE in favor of an equiprobable lottery between BLACK and RED AND WHITE. The latter has a probability  $\frac{1}{2}$  of winning.

Of these two strategies, the first seems simpler, as it merely involves observing that in every lottery in which one chooses BLACK, one need only choose WHITE in an otherwise identical subsequent lottery in order to eliminate ambiguity. The second strategy, by contrast, requires seeing that one needs to pair one's choices in one kind of lottery (betting on a single color) with one's choices in another kind of lottery (with two winning colors).

We do not regard it as plausible that a significant number of subjects employed either of these strategies, because neither strategy seems particularly easy for subjects to grasp. Aside from other considerations, subjects who reason in this way would need to apply some version of the Compound Lottery Axiom that Halevy (2007) has found significant in distinguishing between those subjects classified as ambiguity averse and those who are not.

Notwithstanding our doubts about the likelihood that subjects would form the requisite beliefs and employ one of these strategies, we decided to check whether the low level of ambiguity aversion might nonetheless be partly explained in this way. We therefore ran Version 3 of the experiment, again in the ELSE lab at University

College London. In this version, we eliminated the choices between  $J$  and  $K'$  and  $L$  and  $M'$ , thereby removing the possibility of a subject using the simpler of the ambiguity-eliminating strategies mentioned above. Each round in the previous versions therefore became half as long. In order to keep the number of choices faced by each subject identical to the previous versions, we added two further short rounds, which repeated the first round with different colors. To be precise:

1. As before, in Round 1, the subjects were first navigated through the tree of Fig. 3 to estimate the interval in which to locate the value  $r_1$  that makes a subject indifferent between  $J$  (RED wins) and  $K$  (BLACK wins). Subjects were then navigated through a similar tree to estimate the value  $r_2$  that makes a subject indifferent between  $L$  and  $M$ .
2. Subsequent rounds repeated this round with BLACK replaced by BLUE, YELLOW, and GREEN, respectively.

#### 4 Aggregate results

We had thought that the subjects might learn or otherwise adjust their behavior over time, but the final round data is not significantly different from earlier rounds and so we aggregate the data across all rounds in each version of the experiment. The aggregated results of all three versions of the experiment both for on-site and on-line subjects are given as percentages of the total number of observations in Fig. 6.

Table 1 summarizes our results. A crude criterion in assessing a theory is whether it predicts better than the null hypothesis that subjects answer all questions at random. The first column of the table therefore shows the percentage of times an observation would be made in the long run under this hypothesis. At first sight, all the theories considered seem to pass this test except for the objective maximin criterion (MXN). But how much does the sure-thing principle (STP) explain that is not already explained by the principle of insufficient reason (PIR)? Since STP\PIR does no better than the null hypothesis in the on-site versions of the experiment, the answer would seem to be nothing at all in these versions.

The same reasoning also applies when one asks how much weak or strong ambiguity aversion (WAA or SAA) or the Hurwicz criterion (HWZ with  $h < \frac{1}{2}$ ) explain that is not explained by STP (now interpreted as a measure of approximate ambiguity neutrality). All of HWZ\STP, WAA\STP, and SAA\STP perform no better than the null hypothesis. However, if all the data points in STP are excluded from the population (so that POP is replaced by POP\STP) as in the lower part of the table for 'lax criteria', then each of HWZ\STP, WAA\STP, and SAA\STP performs significantly better than the null hypothesis in on-site Versions 2 and 3 of the experiment. This provides some evidence for ambiguity aversion among subjects who are not approximately ambiguity-neutral.

#### 5 Statistics

This statistical section provides a fuller analysis to address three questions. By how much did the differences in framing between the three versions of our experiment



influence the behavior of the subjects? To what extent is it necessary to invoke ambiguity aversion to explain our data? And, connected to the latter, to what extent does the data offer support for the Hurwicz criterion with a substantial degree of pessimism (with  $h < < \frac{1}{2}$ )?

### 5.1 Kolmogorov–Smirnov test

We answer the preceding questions by appealing to the Kolmogorov–Smirnov (K–S) test, which provides a criterion for deciding whether two samples are generated by the same probability distribution. It is important that the K–S test is non-parametric, because its use shows that some of our data is not normally distributed, which rules out various alternative approaches, including the Chi-Square test.

With one-dimensional data, the K–S statistic  $T$  is obtained by computing the empirical cumulative distribution functions of the two samples to be compared. The value of  $T$  is then the maximum of the absolute difference between these two cumulative distribution functions. Low values of  $T$  indicate that the evidence is not good enough to reject the null hypothesis that the two samples are from the same distribution. To say that the null hypothesis is rejected at the 10% significance level is to say that there is one chance in ten that we are wrong to argue that the two samples do not come from the same distribution.<sup>12</sup>

Lopes et al. (2007) review the problem of applying the K–S test with multi-dimensional data. The problem arises because the manner in which the data points are ordered then becomes significant. Their very severe recommendation requires maximizing over all possible orderings of the data points. Such a procedure seems appropriate when the data is otherwise unstructured, but we exploit the underlying structure of our problem by applying the orthodox one-dimensional K–S test separately to the sums of the columns, the sums of the rows, and the sums of both types of diagonal in each of the  $8 \times 8$  data matrices of Fig. 6. We thereby examine the marginal distributions of  $r_1$ ,  $r_2$ ,  $r_1 + r_2$ , and  $r_1 - r_2$ . The final expression is of particular interest because it can serve as a measure of how far a point  $(r_1, r_2)$  lies from the main diagonal  $r_1 - r_2 = 0$ , where all the data would lie if the subjects were all ambiguity neutral. The K–S test is more reliable when applied to the diagonals than to the rows and columns because the former data is sorted into 15 bins, and the latter into only 8 bins.

### 5.2 Comparing distributions

Table 2 shows K–S statistics for our data.<sup>13</sup> To save on space, we abuse notation by using  $r_1$  and  $r_2$  in this section to refer to our binned data rather than the continuous variables our experiment is intended to estimate.

<sup>12</sup>In computing critical values of  $T$  for the 1%, 5%, and 10% levels, it is necessary to take account of the sample sizes, which vary between different versions of the experiment.

<sup>13</sup>The significance levels  $p$  have been computed from the K–S statistics using the formula  $p = k \times \{(n_1 + n_2)/n_1 n_2\}^{1/2}$ , where  $n_1$  and  $n_2$  are the number of subjects in a population, and  $k = 1.22$  for  $p = 0.1$ ,  $k = 1.36$  for  $p = 0.05$ , and  $k = 1.63$  for  $p = 0.01$ . We therefore treat as independent the means of the four observations obtained from each subject.



**Table 2** Significantly different distributions?

		$r_1$ compared with $r_1$						$r_1 + r_2$ compared with $r_1 + r_2$						
		on-site			on-line			on-site			on-line			
		1	2	3	1	2	3	1	2	3	1	2	3	
on-site	1		0.22	0.21	0.13	0.08	0.11		0.27	0.24	0.14	0.12	0.12	
	2	0.19			0.05	0.23	0.21	0.21	0.07		0.06	0.23	0.21	0.17
	3	0.17	0.05			0.25	0.20	0.21	0.07	0.07		0.18	0.15	0.12
on-line	1	0.11	0.14	0.09			0.06	0.04	0.06	0.06	0.09		0.04	0.08
	2	0.14	0.13	0.09	0.04			0.03	0.05	0.04	0.07	0.03		0.06
	3	0.13	0.13	0.08	0.03	0.06			0.07	0.07	0.10	0.04	0.04	

		$r_2$ compared with $r_2$			$r_1$ compared to $r_2$			$r_1$ compared to $1/2 - r_2$					
		on-site			on-line			on-site			on-line		
		1	2	3	1	2	3	1	2	3	1	2	3
	1	0.18	0.15	0.21	0.05	0.07	0.06	0.27	0.23	0.26	0.05	0.14	0.14
	2												
	3												

Rejection of null hypothesis (distributions the same) at significance level:  
 10%  5%  1%

The top-left tables show Kolmogorov–Smirnov statistics (K–S) that provide a measure of the difference between the marginal distributions of  $r_1$  and  $r_2$  obtained in different treatments. Low values of K–S indicate that there is inadequate evidence to suggest that the distributions are different. The top-right table does the same thing for  $r_1 + r_2$  and  $r_1 - r_2$ . The bottom table compares  $r_1$  and  $r_2$  in the same treatments (which would be the same if the sure-thing principle held) and  $r_1$  and  $\frac{1}{2} - r_2$  (which would be the same if the Hurwicz criterion were to hold)

The data in Table 2 supports the following conclusions.

1. The top-left and top-right tables show that the hypothesis that the on-site data from Version 1 and Version 2 of our experiment come from the same probability distribution is rejected. The hypothesis that the on-site data from Version 1 and Version 3 come from the same distribution is also rejected. Our re-framing of the experiment (and/or the different subject pool) therefore made a substantial difference between Version 1 on the one hand and Versions 2 and 3 on the other.
2. By contrast, the hypothesis that the on-site data from Version 2 and Version 3 of our experiment come from the same probability distribution passes our test. In Section 5.3, we therefore aggregate only the data from Versions 2 and 3. Recall that we have also aggregated the data from all the stages within each version for similar reasons.
3. The first row of the bottom table reports our test of the hypothesis that  $r_1$  and  $r_2$  are drawn from the same distribution, which would be the case if the sure-thing principle held. It shows that in the on-site Versions 1 and 2, one cannot confidently reject this hypothesis. In Version 3, one can reject this hypothesis at the 10% level. This indicates that the data roughly conform to the sure-thing principle, but that there may be other principles determining choice.

4. The second row of the bottom table reports our test of the hypothesis that  $r_1$  and  $\frac{1}{2} - r_2$  are drawn from the same distribution, which would be the case if the Hurwicz principle held. In all on-site versions, there is sufficient evidence to confidently reject the hypothesis. This provides strong evidence against the Hurwicz principle.

### 5.3 Modeling the data

This section fits a simple econometric model to the aggregated data from on-site Versions 2 and 3. The model assumes that subjects basically follow the sure-thing principle. All our data points would then lie on the main diagonal of Fig. 6 if it were not for the further assumption that subjects sometimes diverge from the sure-thing principle when answering the questions in the titration of Fig. 3. To be precise, we assume that the baseline preferences of all subjects satisfy  $r_1 = r_2$ , and that  $r_1$  is normally distributed<sup>14</sup> with mean  $\mu$  and standard deviation  $\sigma$ . When an answer consistent with the baseline preference is in the direction of ambiguity-averse behavior, the model assumes that subjects respond in line with this preference with probability  $a$ , where  $a < 1$ . When an answer consistent with the baseline preference is in the direction of ambiguity-loving behavior, we assume that subjects respond correctly with probability  $d$ , where  $d < 1$ . We therefore have a model with four parameters:  $\mu, \sigma, a, d$ .<sup>15</sup> Can the model be made to fit with no ambiguity aversion ( $a = d$ )? If not, by how much must  $a$  exceed  $d$ ?

To address these questions, we compute two Kolmogorov–Smirnov statistics,  $S$  and  $T$ , using the observed data on the main diagonal for  $S$ , and the sums of data points along parallels to the main diagonal for  $T$ . (The latter exercise is labeled  $r_1 - r_2$  in Table 2.) Low values of  $S$  indicate that the observed data points on the main diagonal of our data matrix are consistent with our model. Low values of  $T$  indicate that deviations from the sure-thing principle are consistent with our model. The 10%, 5%, and 1% significance levels in both cases are 0.10, 0.11, and 0.13. We say that our model is rejected at a particular significance level  $p$  if  $S > p$  or  $T > p$ . (The relevant spreadsheet is available at <http://alkami.org/ells>.)

The results of a hill-climbing exercise in parameter space are as follows. Our model is not rejected at the 5% level when  $\mu = 0.312$ ,  $\sigma = 0.035$ , and  $a = d = 0.82$ . However, it is rejected at the 10% level for all parameter values we examined with  $a = d$ . This means that if one is very careful in rejecting a model (one allows only a 5% chance of wrongly rejecting the model), then a version of our model that does not involve ambiguity aversion passes our test. However, if one is more willing to reject a model (one allows up to a 10% chance of wrongly rejecting it), then no version without ambiguity aversion passes our test.

By contrast, our model cannot be rejected at the 10% level when  $\mu = 0.312$ ,  $\sigma = 0.035$ ,  $a = 0.90$ , and  $d = 0.80$  ( $S = 0.04$ ,  $T = 0.05$ ). The parameter range

<sup>14</sup>Although it makes no difference in practice, we further condition the distribution of  $r_1$  on the requirement that  $0 \leq r_1 \leq \frac{1}{2}$ .

<sup>15</sup>Using different  $a$  and  $d$  for  $r_1$  and  $r_2$  has no significant effect.

within which the latter conclusion can be maintained is small. This means that only a version of our model which posits a modest degree of ambiguity aversion passes the more stringent version of our test.

Notice that the fitted value of  $\mu = 0.312$  is close to  $\frac{1}{3}$  and that  $\sigma = 0.035$  is small. Our assumption that the subjects' basic preferences honor the sure-thing principle is therefore sustained because the data is strongly concentrated in and around the area predicted by the principle of insufficient reason. The analysis shows that we cannot reconcile the model's assumption about the subjects' basic preferences without assuming random deviations from these basic preferences. These deviations need to be biased in the direction of risk-averse behavior, but the degree of bias is small.

For comparison, we also considered the same model with the Hurwicz criterion replacing the sure-thing principle (so that subjects' baseline preferences are assumed normal along the Hurwicz curve of Fig. 2). For  $\mu = 0.350$  and  $\sigma = 0.050$ , the hypothesis that the new model is consistent with the data cannot be rejected at the 5% level when  $a = d = 0.82$  ( $S = 0.04$ , and  $T = 0.11$ ). This result is not very surprising when one notes that the Hurwicz criterion with  $h = \frac{1}{2}$  is indistinguishable from the principle of insufficient reason.

In summary, our model best fits the data when it describes our population as if:

- (i) Each subject's baseline preferences approximate the principle of insufficient reason;
- (ii) Each subject has a modest tendency to randomly diverge from these baseline preferences, with diversion in an ambiguity-averse direction being somewhat more likely.

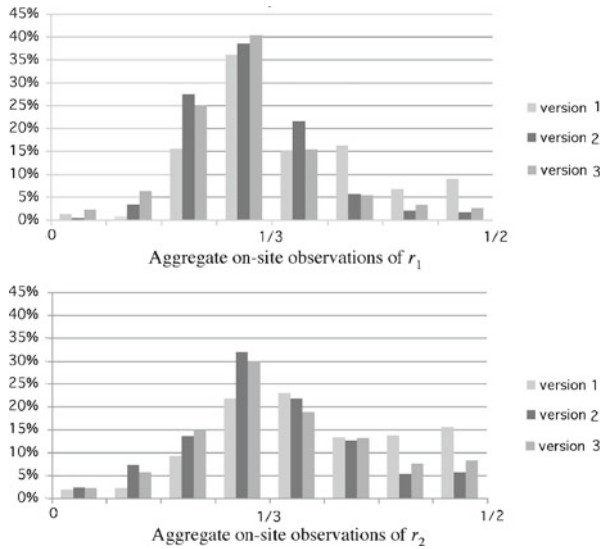
Indeed, this version of our model cannot confidently be rejected. Our modeling exercise therefore supports a modest degree of ambiguity aversion, but refutes versions of the Hurwicz criterion with  $h$  significantly less than  $\frac{1}{2}$ .

## 6 Discussion

1. The principle of insufficient reason has the most predictive power in our experiment. About one third of our observations lie in the shaded region corresponding to PIR in Fig. 4.
2. Theories that postulate a large level of ambiguity aversion all perform badly compared with PIR.<sup>16</sup> The objective maximin criterion performs very badly indeed.<sup>17</sup> However, as evidenced by the lower part of Table 1 and our model, the data do provide evidence for a modest degree of ambiguity aversion.

<sup>16</sup>Recent working papers by Ahn et al. (2010) and Charness et al. (2012) report similar results with a different experimental design.

<sup>17</sup>The same goes for the minimax regret criterion, since this coincides with the maximin criterion in our experiment.



**Fig. 7** Histograms of aggregate on-site observations. Our criterion for *strong* ambiguity requires both  $r_1 < \frac{1}{3}$  and  $r_2 > \frac{1}{3}$ . If attention were restricted to  $r_1$ , more subjects in all three versions would count as strongly ambiguity averse than we report (53%, 68%, 73%). The effect is considerably weaker for  $r_2$  (65%, 46%, 48%). Note also the significant alteration in behavior between Versions 1 and 2 in which subjects were offered exactly the same problem but framed differently

3. Why do we not observe as much ambiguity aversion as is often reported? One possible reason is that our experimental protocol reduces suspicion of deceit on the part of the experimenter. A second potential reason is that Versions 1 and 2 of our experiment allow sophisticated subjects to treat all probabilities as objective. However, levels of ambiguity aversion remain slight in Version 3, where it is harder to pull off the same trick. A third reason is that our ambiguity aversion criteria are more demanding than in some studies, a point that we take up in the next item.
4. Our criteria for ambiguity aversion require that a subject be consistently ambiguity averse in two different (but related) comparisons. By contrast, some experiments that report high levels of ambiguity aversion require a preference for a risky over an ambiguous option in only one comparison.<sup>18</sup> But it is presumably uncontroversial that ambiguity aversion judgments need to be reliable to be a useful predictive tool in real-life situations. Figure 7 illustrates that our more demanding criteria make a difference. If one were to pay attention only to estimates of  $r_1$ , then one would find what would seem to be substantial evidence for strong ambiguity aversion, especially in the case of Version 3 (where 73% of observations are consistent with strong ambiguity aversion as opposed to the 60–70% commonly reported).
5. Figure 7 also shows that part of the reason that our criteria make a difference is that there is more ambiguity aversion in our one-winning-color choices (for which the indifference probability is  $r_1$ ) than in our two-winning-color choices

(for which the indifference probability is  $r_2$ ). Our tentative conjecture is that to some subjects, the ambiguous option is easier to discern in the one-winning-color case than in the two-winning-color case. In the one-winning-color case, the ambiguity present in betting on BLACK is marked, and the lack of ambiguity in betting on RED is obvious. By contrast, in the two-winning-color case, the ambiguity present in betting on RED AND WHITE may be less marked to some (because subjects know the probability of part of the winning combination) and the lack of ambiguity in betting on BLACK AND WHITE may not be obvious to some (because it requires observing that while the individual component colors are ambiguous, the compound is not).

6. The robustness issue also arises insofar as our experiment provides another example of the kind of framing effects emphasized by Kahneman and Tversky (1981). In particular, subjects responded differently in Versions 1 and 2 of the experiment, even though the questions they were asked were logically identical.
7. Another potential explanation for the different pattern of response between Versions 1 and 2 also bears on this issue. These versions were conducted in different laboratories with different subject pools. Version 1 was conducted in Harvard with a pool of largely American subjects of whom about one third were students. Version 2 was conducted at University College London with a pool of largely British subjects of whom about two thirds were students. The mean age of the British pool was about seven years younger than the American pool.<sup>19</sup> Though we therefore cannot be certain whether the difference is due to framing effects or the composition of the subject pool, either way, this shift casts doubt on the robustness of subjects' pattern of response in conditions of uncertainty.

## 7 Psychological and demographic correlations

Psychologists define optimism and pessimism as positive and negative outcome expectancies, and it has been proposed that people with a predisposition to expecting things to turn out well might perceive an ambiguous situation differently from those expecting things to turn out badly. Previous work has indicated an inverse relationship between ambiguity aversion and pessimism within the paradigm of the Ellsberg paradox (Pulford 2009; Lauriola et al. 2007). We therefore explored this relationship

---

<sup>18</sup>For example, Keren and Gerritsen (1999) ask subjects to choose between betting on a red ball drawn from an urn which has a known probability of  $\frac{1}{3}$  of yielding a red ball and betting on a green ball or on a blue ball drawn from a different, ambiguous urn, in which green and blue together make up slightly more than  $\frac{2}{3}$  of the balls. They conclude that a large majority preference for betting on red is evidence of ambiguity aversion. Liu and Colman (2009) similarly use a one-choice setup. They offer subjects a choice between betting on a red ball drawn from an urn with a known probability of  $\frac{1}{2}$  of yielding a red ball and betting on a red ball drawn from an urn containing red and green balls in an unknown proportion. The prize from the uncertain urn was somewhat larger than the prize from the risky urn. They take a majority preference to bet on red from the risky urn to be evidence of ambiguity aversion.

<sup>19</sup>Genders were roughly equal in each case.

measuring the degree of pessimism/optimism with the Life Orientation Test, Revised (LOT-R) of Scheier et al. (1994). A subject's ambiguity aversion was measured by averaging  $r_2 - r_1$  for the subject's four choice sets, with a negative score indicating ambiguity-loving behavior, and a positive result indicating ambiguity-averse behavior. These averaged scores ranged from  $-0.41$  to  $0.46$  with a mean of  $0.011$  (.081). LOT-R scores ranged from 0 (most pessimistic) to 24 (most optimistic), with a mean of  $14.5$  (5.0).<sup>20</sup>

We compared these measures of ambiguity aversion and pessimism/optimism for each version separately for on-line and on-site participants using the Kruskal–Wallis test, Levenes' test for equality of variances, independent samples  $t$ -tests, and scaled JZS Bayes Factors (Rouder et al. 2009). Our measure of pessimism/optimism was not associated with ambiguity aversion.

Some research suggests that ambiguity aversion differs across genders (Borghans et al. 2009; Powell and Ansic 1997). However, no gender differences in mean ambiguity aversion were identified.

Comparisons of our measure of ambiguity aversion with other personal and demographic variables also showed no robust correlation for any variable. A full report of these comparative analyses is available at <http://alkami.org/ells>.

## 8 Conclusion

Ambiguity aversion has been regularly observed in a majority of subjects' choices in the standard Ellsberg experiment. It is also regarded as a prevailing phenomenon for choices involving gains generally. We designed a new experiment to examine how much of this phenomenon could be explained by behavior in accordance with the Hurwicz criterion. However, ambiguity aversion was less pronounced than in many other studies and we found little evidence of behavior in accordance with the Hurwicz criterion. Indeed, the Hurwicz criterion with a substantial degree of pessimism is clearly inconsistent with our findings. Only the principle of insufficient reason had marked predictive power in our experiment. We twice changed the framing of our experiment, which had a significant effect on some features of the subjects' behavior, but our basic conclusions were left unchanged.

We tentatively attribute our findings to two aspects of our study. First, we worked hard to eliminate suspicion that the experimenter might be manipulating the gambles. Second, our criteria for ambiguity aversion are more demanding than in some studies because they require a subject to display aversion to the ambiguous option in *two* different but related types of choices. Subjects whose choices do not meet such strict criteria are not robustly ambiguity averse.

---

<sup>20</sup>These LOT-R results are consistent with the norms reported in Scheier et al. (1994).

## References

- Abdellaoui, M., Baillon, A., Placido, L., Wakker, P. (2011). The rich domain of uncertainty: source functions and their experimental implementation. *American Economic Review*, *101*, 695–723.
- Ahn, D., Choi, S., Gale, D., Kariv, S. (2010). Estimating ambiguity aversion in a portfolio choice experiment. <http://www.econ.nyu.edu/user/galed/papers.html>. Accessed January 2011.
- Arrow, K., & Hurwicz, L. (1972). An optimality criterion for decision making under ignorance. In C. Carter, & J. Ford (Eds.), *Uncertainty and expectations in economics*. Oxford: Basil Blackwell.
- Binmore, K. (2009). *Rational decisions*. Princeton: Princeton University Press.
- Borghans, L., Heckman, J., Golsteyn, B., Meijers, H. (2009). Gender differences in risk and ambiguity aversion. *Journal of the European Economic Association*, *7*, 649–658.
- Brewer, K. (1965). Decisions under uncertainty: comment. *Quarterly Journal of Economics*, *79*, 657–653.
- Camerer, C., & Weber, M. (1992). Recent developments in measuring preferences: uncertainty and ambiguity. *Journal of Risk and Uncertainty*, *5*, 325–370.
- Charness, G., Karni, E., Levin, D. (2012). Ambiguity attitudes: An experimental investigation. <http://ideas.repec.org/p/jhu/papers/590.html>.
- Chow, C., & Sarin, R. (2001). Comparative ignorance and ambiguity aversion. *Journal of Risk and Uncertainty*, *22*, 129–139.
- Curley, S., & Yates, J. (1989). An empirical evaluation of descriptive models of ambiguity reactions in choice situations. *Journal of Mathematical Psychology*, *33*, 397–427.
- Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms. *Quarterly Journal of Economics*, *75*, 643–669.
- Ellsberg, D. (2001). *Risk, ambiguity, and decision*. New York and London: Garland Publishing.
- Etner, J., Jelewa, M., Tallon, J.-M. (2012). Decision theory under ambiguity. *Journal of Economic Surveys*, *26*, 324–270.
- Fox, C., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *Quarterly Journal of Economics*, *110*, 585–603.
- Ghirardato, P., & Marinacci, M. (2002). Ambiguity made precise: a comparative foundation. *Journal of Economic Theory*, *102*, 251–289.
- Gilboa, I. (2004). *Uncertainty in economic theory: Essays in honor of David Schmeidler's 65th birthday*. London: Routledge.
- Gilboa, I., & Schmeidler, D. (1989). Maximin expected utility with non-unique prior. *Journal of Mathematical Economics*, *18*, 141–153.
- Halevy, Y. (2007). Ellsberg revisited: an experimental study. *Econometrica*, *75*, 503–506.
- Hey, J., Lotito, G., Maffioletti, A. (2010). The descriptive and predictive adequacy of decision making under uncertainty/ambiguity. *Journal of Risk and Uncertainty*, *41*, 81–111.
- Hogarth, R., & Einhorn, H. (1990). Venture theory: a model of decision weights. *Management Science*, *36*, 780–803.
- Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., Camerer, C. (2005). Neural systems responding to degrees of uncertainty in human decision making. *Science*, *310*, 1680–1683.
- Hurwicz, L. (1951). *Optimality criteria for decision making under ignorance (Vol. 370)*. Cowles Commission Discussion Paper, Statistics.
- Kadane, J. (1992). Healthy scepticism as an expected-utility explanation of the phenomena of Allais and Ellsberg. In J. Geweke (Ed.), *Decision-making under risk and uncertainty: New models and empirical findings*. Dordrecht: Kluwer.
- Kahn, B., & Sarin, R. (1988). Modeling ambiguity in decisions under uncertainty. *Journal of Consumer Research*, *15*, 265–272.
- Kahneman, D., & Tversky, A. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458.
- Keren, G., & Gerritsen, L. (1999). On the robustness and possible accounts of ambiguity aversion. *Acta Psychologica*, *103*, 149–172.
- Lauriola, M., Levin, I., Hart, S. (2007). Common and distinct factors in decision making under ambiguity and risk: a psychometric study of individual differences. *Organizational Behavior and Human Decision Processes*, *104*, 130–149.
- Lopes, R., Reid, I., Hobbes, P. (2007). The Two-Dimensional Kolmogorov–Smirnov Test XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research, Amsterdam. (Published by Proceedings of Science).

- Luce, R., & Raiffa, H. (1957). *Games and decisions*. New York: Wiley.
- Liu, H.-H., & Colman, A. (2009). Ambiguity aversion in the long run: repeated decisions under risk and uncertainty. *Journal of Economic Psychology*, 30, 277–284.
- MacCrimmon, K., & Larsson, S. (1979). Utility theory: Axioms versus ‘paradoxes’. In M. Allais, & O. Hagen (Eds.), *Expected utility and the Allais paradox* (pp. 333–409). Dordrecht: Reidel.
- Milnor, J. (1954). Games against nature. In R. Thrall, C. Coombs, R. Davies (Eds.), *Decision processes*. New York: Wiley.
- Powell, M., & Ansic, D. (1997). Gender differences in risk behavior in financial decision-making. *Journal of Economic Psychology*, 18, 605–828.
- Pulford, B. (2009). Is luck on my side? Optimism, pessimism, and ambiguity aversion. *Quarterly Journal of Experimental Psychology*, 62, 1079–1087.
- Raiffa, H. (1961). Risk, ambiguity, and the Savage axioms: comment. *Quarterly Journal of Economics*, 75, 680–694.
- Rouder, J., Speckman, P., Sun, D., Morley, R., Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, 62, 1079–1087.
- Savage, L. (1954). *The foundations of statistics*. New York: Wiley.
- Scheier, M., Carver, F., Bridges, M. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a re-evaluation of the life orientation test. *Journal of Personality and Social Psychology*, 67, 1063–1078.
- Schneeweiss, H. (1968). Spieltheoretische analyse des Ellsberg-paradoxons. *Zeitschrift für die Gesamte Staatswissenschaft*, 124, 249–255.
- Trautmann, T., Vieider, F., Wakker, P. (2008). Causes of ambiguity aversion: known versus unknown preferences. *Journal of Risk and Uncertainty*, 36, 225–243.
- Viscusi, W.K., & Magat, W. (1992). Bayesian decisions with ambiguous belief aversion. *Journal of Risk and Uncertainty*, 5, 371–378.
- Viscusi, W.K. (1997). Alarmist decisions with divergent risk information. *Economic Journal*, 107, 1657–1670.
- Wakker, P. (2010). *Prospect theory for risk and ambiguity*. Cambridge: Cambridge University Press.
- Wald, A. (1950). *Statistical decision theory*. New York: Wiley.