An Experiment on the Ellsberg Paradox

Ken Binmore
Economics Department
University College London
London WC1E 6BT, UK

Lisa Stewart
Psychology Department
Australian National University
Canberra ACT 0200, Australia

Alex Voorhoeve (corresponding author)
Department of Philosophy, Logic and Scientific Method
London School of Economics
London WC2E 2AE, UK
tel: +44-20-7955-6579
fax: +44-20-7955-6845
email: a.e.voorhoeve@lse.ac.uk

Abstract: Experimental results on the Ellsberg paradox typically reveal behavior that is commonly interpreted as ambiguity aversion. The experiments reported in the current paper were intended to examine the extent to which such ambiguity aversion can be captured by the Hurwicz criterion. However, contrary to our expectations, Laplace's principle 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, maxmin criterion, principle of insufficient reason.

JEL classification: C91, D03.

Even bonobos like those pictured above from the Lola Ya Sanctuary are reported to be ambiguity averse in an experiment on the Ellsberg paradox:

http://www.sciencedaily.com/releases/2010/11/101129203344.htm

An Experiment on the Ellsberg Paradox

Ken Binmore, Lisa Stewart and Alex Voorhoeve[1]

1 Ellsberg Paradox

Daniel Ellsberg [11, 12] famously proposed an experiment whose much replicated results have become known as the Ellsberg paradox because they are inconsistent with the predictions of expected utility theory.[2]

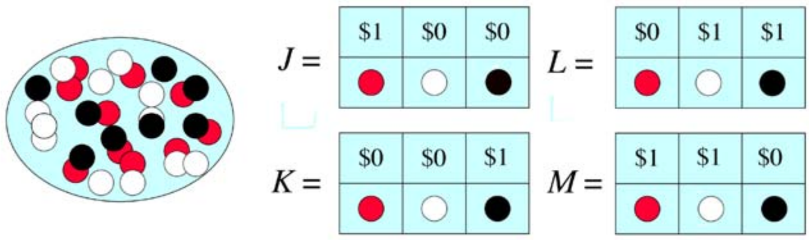

Figure 1: Ellsberg paradox: In the version illustrated an urn contains ten red balls and another twenty balls of which it is only known that they are either black or white. A ball is chosen at random from the urn, the color of which determines the award of a prize (which is here taken to be one dollar).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

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 Figure 1 represent various reward

[1] Parts of the experimental data were gathered while Lisa Stewart was a researcher in the Harvard Psychology Department and Alex Voorhoeve was Faculty Fellow at Harvard's Safra Center for Ethics. We thank the Decision Science Laboratory at Harvard's Kennedy School of Government and the ELSE laboratory at University College London for the generous use of their facilities, and the Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD) for financial support. Ken Binmore thanks both the British Economic and Social Research Council through the Centre for Economic Learning and Social Evolution (ELSE) and the British Arts and Humanities Research Council through grant AH/F017502. Alex Voorhoeve thanks the Safra Center for Ethics for its Faculty Fellowship. Preliminary results were presented at Harvard University, the LSE, and the University of Siena. We thank Richard Bradley, Joshua Greene, Glenn Harrison, Jimmy Martinez, Joe Swierzbinski, and those present at our seminars for their comments.

[2] Ahn *et al* [1], Becker and Browson [2], Brewer [6], Camerer and Weber [7], Chow and Sarin [8], Curley and Yates [10], Fellner [14], Etner *et al* [13], Fox and Tversky [15], Frisch and Baron [16], Halevy [19], Hey *et al* [20], Hsu *et al* [21], Keren and Gerritsen [25], Liu and Colman [27], Pulford [30], Trautman *at al* [36]. The link http://aversion-to-ambiguity.behaviouralfinance.net/ leads to many more papers on ambiguity aversion published since the year 2000.

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.

Ambiguity aversion. The standard 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 $R + B$ that might be anything in the range $[\frac{1}{3}, 1]$.

Such ambiguity aversion is often said to be a "powerful and robust" effect, as in Lui and Colman [27]. However, the partial survey of the experimental literature offered by Etner *et al* [13] suggests that the question is more open than such language suggests, and that 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.

A subject who honors Savage's [32, p. 21] sure-thing principle will have

$$r_1 = r_2. \tag{1}$$

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.

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. However, we were surprised to discover that Laplace's principle of insufficient reason holds up better than rival predictions in the context of our experiment, in spite of our attempt to challenge the phenomenon by altering the framing of the experiment. A recent working paper by Ahn *et al* [1] reports similar results with a different experimental design.

2 Theories

Figure 2 illustrates various alternative theories of choice behavior under uncertainty. These all have 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 [17]). Our case is simpler because it leaves no room for 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.

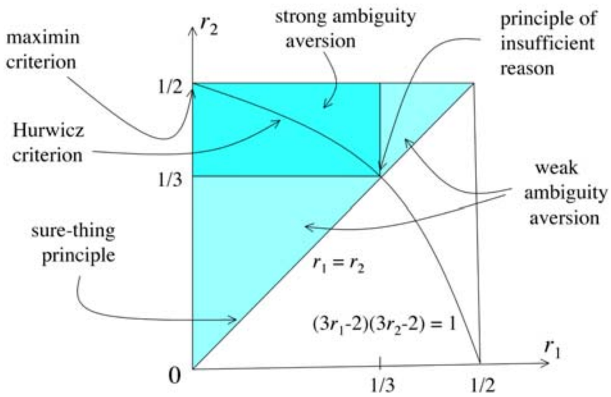

Figure 2: Theoretical possibilities: Bayesian decision theory predicts an outcome on the diagonal of the square. Laplace's principle of insufficient reason says that the probabilities of a black or a white ball should be taken to be equal, as there is no reason to suppose one more likely than the other.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

Maximin criterion (MXN). The most widely discussed alternative to Bayesian decision theory was proposed by Wald [37] and has been developed since that time by numerous authors (see Gilboa [18]). 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. Applying this 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.[3]

Hurwicz criterion (HWZ). Less attention is paid to a proposal of Hurwicz [22], which was axiomatized by Milnor [28]. It balances the pessimism of the maximin criterion against the optimism of what might be called the maximax criterion. The 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 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 maximin criterion. The case $h = 1$ corresponds to the 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[4]

$$(3r_1 - 2)(3r_2 - 2) = 1\,. \tag{3}$$

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 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 actually 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.

[3]The 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 [32] 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 [32, 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".)

[4]Binmore [3, p. 166] offers reasons for replacing the weighted arithmetic mean of the Hurwicz criterion by a weighted geometric mean. The change is not relevant for the range of money payments considered here because, to a first order of approximation, it also yields equation (3). 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.

Laplace's principle of insufficient reason (PIR). This principle says that events should be assigned the same 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

$$r_1 = r_2 = \tfrac{1}{3} \,. \tag{4}$$

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 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

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 Figure 3.[5]

Such a 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 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.[6]

[5] In this exercise, it is assumed that $K \succ J$ when $R = 0$, and $J \succeq K$ when $R \geq \frac{1}{2}$; also $L \succ M$ when $R = 0$, and $M \succeq L$ when $R \geq \frac{1}{2}$.

[6] 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

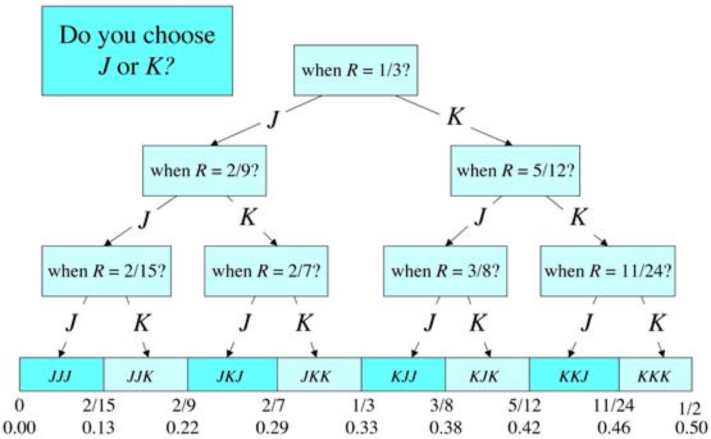

Figure 3: A titration: The tree shows the questions asked about what happens when the probability of RED is R in order to locate r_1 in one of eight subintervals of $[0, \frac{1}{2}]$. For example, someone who answers KJJ is assigned a value of r_1 satisfying $\frac{1}{3} \leq r_1 \leq \frac{3}{8}$. 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.

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 would be indifferent between J and K (and 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, do we not shrink PIR to obtain a smaller region pir. The region nstp in Figure 4 should be thought of only as the neutral part of 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

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 misjudgement at the first question but answered later questions accurately).

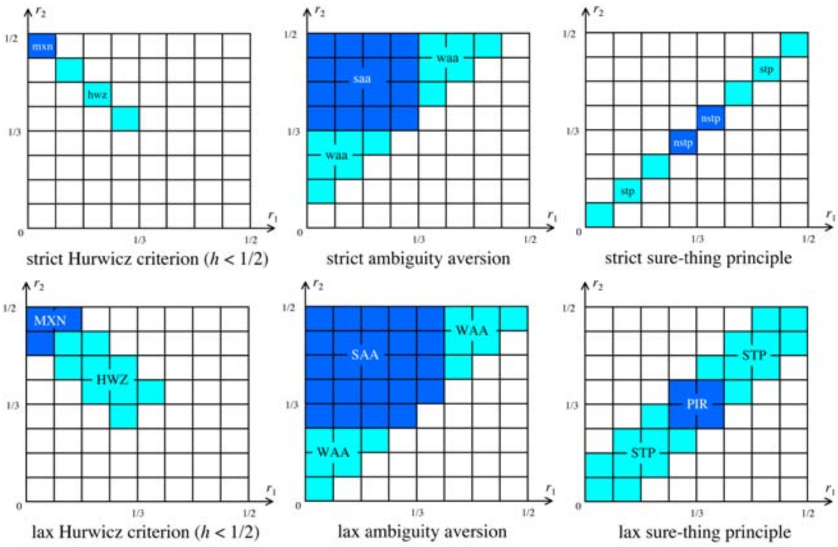

Figure 4: Regions of interest: The results are reported using chessboards showing the percentage of times that (r_1, r_2) is observed in one of 64 possible squares. The chessboards in the top row of the figure correspond to a strict interpretation of each region, which allows no margin for mistaken choices by the subjects. (The regions of interest differ somewhat from Figure 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.

(PIR). In considering such issues, we use the notation WAA\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 were asked to choose whether to bet on J or K for various values of the probability R of drawing a red card, and to choose whether to bet on L or M for various values of the probability R of drawing a red card. For example, Figure 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:

http://www.alkami.org/ells/

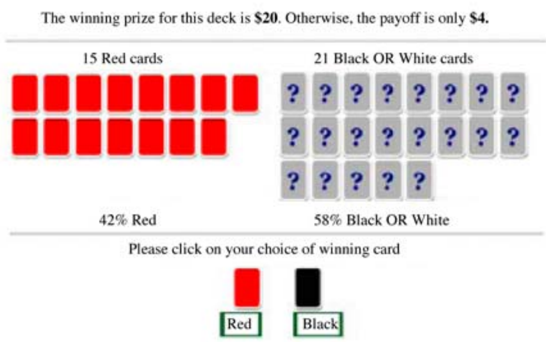

Figure 5: Experimental interface: When confronted with this particular interface, the subject is being asked whether they prefer J or K when the probability of drawing a red card is $R = \frac{5}{12}$.

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 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* [20], 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 practise choices, subjects were invited to the front of the room to witness one of the practise 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 practise 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 every subject had tailor-made bets constructed and played for him or her.)

This procedure may be relevant to the relatively low levels of ambiguity aversion we observed. For example, Pulford [30] reports significantly more ambiguity aversion after drawing attention to the possibility of experimental deceit. However, 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 [6], Schneeweiss [34], Kadane [23]). 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 [34]). Researchers are then at risk of misinterpreting such play as the use of an ambiguity-averse strategy in a one-person decision problem.[7]

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 to check up on how the cards were shuffled and dealt).[8] The behavior of on-line subjects turned out to resemble that of on-site subjects, but is much noisier. Our commentary therefore concentrates on the on-site results.

[7] Savage [32, 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.

[8] 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://rsvp.alkami.org). Participants recruited via Mechanical Turk were paid a token $0.05 for completing the approximately six-minute study.

3.1 Version 1

Following the practise 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. For this purpose, in Round 1 of the experiment, subjects were faced with choices between J and K (to reveal the interval for r_1) and L and M (to reveal the interval for r_2), as well as between J and K' (which is like K, except that WHITE replaces BLACK as the winning color) and between L and M' (which is like M, except that the winning colors are RED and BLACK).

The experiment consisted of two main rounds, some brief questionnaires, the actual play of two randomly selected gambles, and payment of the subjects:

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.

(a) Three choices between J (RED wins) and K (BLACK wins) with values of R (the probability of RED) varying according to Figure 3.

(b) Three choices between J and K' (WHITE wins) with values of R varying according to Figure 3.

(c) Three choices between L (WHITE and BLACK win) and M (RED and WHITE win) with values of R varying according to Figure 3. (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) with values of R varying according to Figure 3. (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 Figure 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. After both rounds were over, subjects were taken one-by-one through the questions from a Life Orientation Test as revised by Scheier *et al* [33].

4. Subjects were next asked five questions about whether the tasks and questions had been clear, and whether any problems arose during the experiment.

5. Finally, two gambles (one from each round) were chosen at random by the computer for each subject to be played for real. Each subject saw these two gambles and his or her choices in these gambles displayed on his or her screen. The appropriate decks of cards were assembled and shuffled. Payment was made on the basis of the winning card(s) in the shuffled decks. (The prizes were chosen to render subjects choosing according to the principle of insufficient reason indifferent about which

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

3.2 Version 2

The data from Version 1 are displayed in Figure 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 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.

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 (including the pilot which did not lead to changes in the new design).[9]

Though, as our analysis in Section 6 shows, the results of the on-site subjects for Version 2 differed significantly in some respects from the results of the on-site 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. One possible explanation is that on-site subjects could theoretically convert the problem into one in which an unambiguous probability 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.

[9] The prizes were approximately equal to their previous dollar values but were denominated in British pounds. The average payout was around £13.

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 only 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 [19] 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 Figure 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.

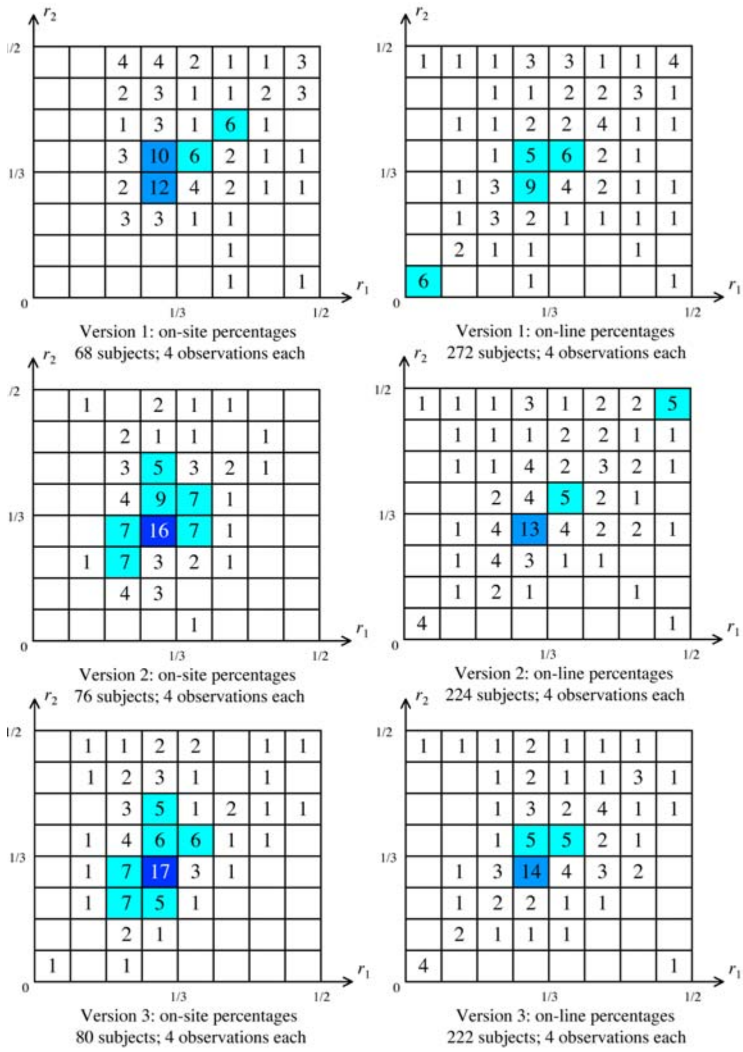

Figure 6: Summary of aggregate results. The on-site and on-line data for all observations for each version of the experiment are shown. Shaded squares indicate a particularly high concentration of responses.

4 Aggregate Results

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 Figure 6. We had thought that the subjects might adjust their behavior over time, but the final round data is not significantly different from earlier rounds and so we do not report this.

lax criteria

	0	aggregate percentages for on-site versions 1	2	3	aggregate percentages for on-line versions 1	2	3
POP	100	100	100	100	100	100	100
MXN	5	1	1	1	2	2	2
HWZ (h<1/2)	19	38	48	45	28	32	33
SAA	38	56	63	64	45	46	47
WAA	56	74	75	80	72	73	73
STP	34	60	69	65	62	64	66
PIR	6	31	39	32	24	25	28
HWZ\STP	14	11	15	16	8	10	9
SAA\STP	30	25	20	27	20	20	18
WAA\STP	33	27	22	27	22	22	19
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	21	27	49	45	22	28	27
SAA\STP	45	64	65	77	54	54	52
WAA\STP	50	68	69	78	58	61	56

strict criteria

	0	aggregate percentages for on-site versions 1	2	3	aggregate percentages for on-line versions 1	2	3
pop	100	100	100	100	100	100	100
mxn	2	0	0	0	1	1	1
hwz (h<1/2)	6	13	13	9	8	6	7
saa	25	32	27	28	20	20	19
waa	44	42	42	43	36	36	33
stp	13	32	33	37	36	37	40
nspt	3	18	23	23	15	18	19
stp\nspt	9	15	10	14	21	19	22

significant deviation at the 5% level from the null hypothesis toward the criterion

significant deviation at the 1% level from the null hypothesis toward the criterion

Table 1: Summary of aggregate percentages. The acronyms for different theories appear in Section 2. The upper part of the table for lax criteria shows percentages for each region of Figure 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.)

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 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 all of HWZ\STP, WAA\STP and SAA\STP perform significantly better than the null hypothesis in on-site versions 2 and 3 of the experiment.

5 General Discussion

This section reviews our results.

1. The data from on-line subjects is markedly noisier than that of on-site subjects. (The hypothesis that on-line subjects paid less attention to their choice problem than on-site subjects is supported by the fact that at 2.2 seconds, the mean response time of on-line subjects to each choice was just over half the mean response times of on-site subjects.) The discussion that follows therefore focuses on the on-site data (in which the subjects were paid and able to monitor the honesty of the shuffling and dealing of the cards).

2. The theory that has the most predictive power in our experiment is Laplace's principle of insufficient reason (which we cannot distinguish from the Hurwicz criterion when $h \approx \frac{1}{2}$). About one third of our observations lie in the shaded region corresponding to PIR in Figure 4.[10]

3. Theories that postulate a marked level of ambiguity aversion all perform badly compared with PIR; the maximin criterion performs very badly indeed.[11]

4. What of the observations left after those counted as consistent with ambiguity neutrality (STP) are excluded? When statistical tests are carried out with the whole population POP replaced by POP\STP, it is possible to reject the null hypothesis that the data for SAA\STP and WAA\STP was generated by choices made at random at the 5% level (1% in Version 3). The Hurwicz criterion (HWZ\STP) survives this test at the 1% level for both versions 2 and 3, although the maximin criterion continues to perform very badly.

[10] The modal square in all versions of the experiment is consistent with a small deviation from the strict principle of insufficient reason towards ambiguity aversion when comparing J and K and away from ambiguity aversion when comparing L and M.

[11] The same goes for the minimax regret criterion, since this coincides with the maximin criterion in our experiment.

5. How come we do 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 much harder to pull off the same trick. A third reason is that our ambiguity aversion criteria are more demanding than is usual, a point that we take up in the next item.

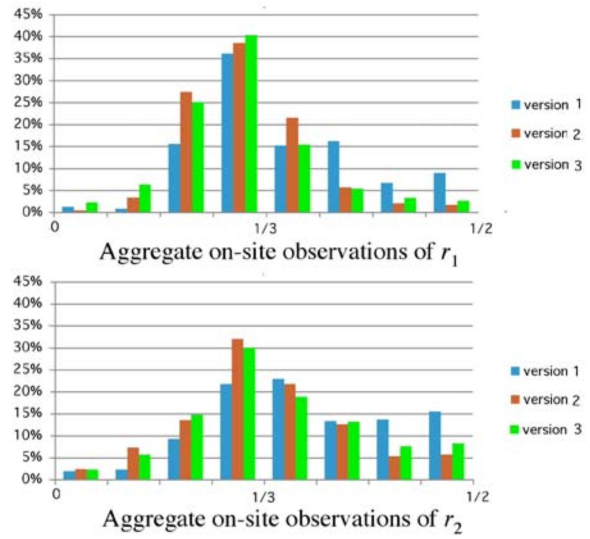

Figure 7: Histograms of aggregate on-site observations. Our criterion for strong ambiguity requires both $r_1 < \frac{1}{3}$ and $r_2 > \frac{1}{3}$. If attention were restricted to r_1, more subjects in all three versions would count as ambiguity averse than we report (53%, 68%, 73%). The effect is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

6. Our criteria for ambiguity aversion require that a subject be consistently ambiguity averse in two decision problems, whereas most experiments only look at one type of decision problem. Figure 7 illustrates this point. If one were to pay attention only to estimates of r_1, then one would find what would seem to be strong evidence for ambiguity aversion, especially in the case of Version 3 (where 74% of observations are consistent with ambiguity aversion as opposed to the 60–70% commonly reported.) But it is presumably uncontroversial that ambiguity aversion needs to be a reliable phenomenon for it to be useful as a predictive tool in real-life applications.

7. The robustness issue also arises insofar as our experiment provides another

example of the kind of strong framing effects emphasized by Kahneman and Tversky [24]. In particular, subjects responded differently in versions 1 and 2 of the experiment, even though the questions they were asked were logically identical (Table 2).

8. 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.[12] 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.

6 Supporting Statistics

The statistical results reported here should be regarded as suggestive only, because the data of Figure 6 is a two-dimensional array in which each dimension has been sorted into only eight bins by the titration of Figure 3. We nevertheless employ one-dimensional Kolmogorov-Smirnov (K-S) tests since no other off-the-peg statistical technique seems better suited to our data.[13] 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.

6.1 Comparing Distributions

Recall that the Kolmogorov-Smirnov test can be used to judge whether two empirical samples have been generated by the same probability distribution. Low values of the K-S statistic (which is the maximum absolute difference between the empirical cumulative distribution functions) suggest that the evidence is not good enough to reject the null hypothesis that the two samples are from the same distribution.

Table 2 lists Kolmogorov-Smirnov statistics in the cases that seem of interest.[14] The top-left tables examine the marginal distributions of r_1 and r_2, thereby struc-

[12] Genders were roughly equal in each case.

[13] The Chi-Square test applies to normally distributed data, but our data mostly fails a test for normality. Two-dimensional versions of the K-S test assume that no underlying structure is available to order our 64 bins.

[14] 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.

r_1 compared with r_1 (upper triangle); r_2 compared with r_2 (lower triangle)

		on-site 1	on-site 2	on-site 3	on-line 1	on-line 2	on-line 3
on-site	1		0.22	0.21	0.13	0.08	0.11
	2	0.19		0.05	0.23	0.21	0.21
	3	0.17	0.05		0.25	0.20	0.21
on-line	1	0.11	0.14	0.09		0.06	0.04
	2	0.14	0.13	0.09	0.04		0.03
	3	0.13	0.13	0.08	0.03	0.06	

$r_1 + r_2$ compared with $r_1 + r_2$ (upper triangle); $r_1 - r_2$ compared with $r_1 - r_2$ (lower triangle)

		on-site 1	on-site 2	on-site 3	on-line 1	on-line 2	on-line 3
on-site	1		0.27	0.24	0.14	0.12	0.12
	2	0.07		0.06	0.23	0.21	0.17
	3	0.07	0.07		0.18	0.15	0.12
on-line	1	0.06	0.06	0.09		0.04	0.08
	2	0.05	0.04	0.07	0.03		0.06
	3	0.07	0.07	0.10	0.04	0.04	

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

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

Table 2: Significantly different distributions? 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).

turing our 8×8 data matrix in the N-S and E-W directions. The top-right table structures our data in the NW-SE (r_1+r_2) and NE-SW directions (r_1-r_2), a choice motivated by our interest in the Hurwicz criterion and the sure-thing principle (see Figure 4). In these tables, we therefore summarize the data using 15 bins containing the sums of the data points along parallels to the diagonals of the data matrix. The bottom table compares r_1 and r_2 (which should have the same distribution if the sure-thing principle applies), and r_1 and r_2 with its ordering reversed (which should have the same distribution if the Hurwicz criterion applies).

The following conclusions tally with the General Discussion of Section 5:

1. Our change in framing from version 1 to version 2 did lead to a change in the behavior of on-site subjects, even though the decision problem they faced was theoretically identical in both cases.

2. Evidence of a corresponding change for on-site subjects in moving from version 2 to version 3 is lacking.

3. The Hurwicz criterion performs badly for on-site subjects.

4. On-line subjects seem relatively insensitive to our framing changes. (Recall that totally random samples would be classified as coming from the same distribution.) There is evidence that on-site subjects in versions 2 and 3 behave significantly differently from on-line subjects, which would seem to be bad news for experiments in this area that rely only on on-line subjects.

6.2 Modeling the Data

As an exercise in purely descriptive statistics, this section fits a simple model to the aggregated data from on-site versions 2 and 3 (Table 2). The model assumes that subjects basically follow the sure-thing principle. All data would then lie on the main diagonal of Figure 6 if it were not for the further assumption that subjects sometimes diverge from their "true" preferences when answering the questions in the titration of Figure 3. To be precise, we assume that the true preferences of all subjects satisfy $r_1 = r_2$, and that r_1 is normally distributed[15] with mean μ and standard deviation σ. When a true answer is in the direction of ambiguity-averse behavior, we assume that subjects respond correctly only with probability a, where $a < 1$. When a true answer is in the direction of ambiguity-loving behavior, we assume that subjects respond correctly only with probability d, where $d < 1$. We therefore have a model with four parameters: μ, σ, a, d.[16] 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 suggest that the observed data points on the main diagonal of our data matrix are consistent with our model. Low values of T suggest 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 the previously referenced website.)

The results of a hill-climbing exercise in parameter space are easily summarized. Our model is rejected at the 10% level for all parameter values we examined with $a = d$. On the other hand, the hypothesis that our model is consistent with the data when $\mu = 0.312$, $\sigma = 0.035$, $a = 0.90$, and $d = 0.80$ cannot be rejected at the 10% level ($S = 0.04$, $T = 0.05$).

For comparison, we also considered the same model with the Hurwicz criterion replacing the sure-thing principle (so that subjects' "true" preferences are assumed normal along the Hurwicz curve of Figure 2). For appropriate values of μ and σ,[17]

[15] Although it makes no difference in practise, we further condition the distribution of r_1 on the requirement that $0 \le r_1 \le \frac{1}{2}$.

[16] Using different error probabilities for r_1 and r_2 has no significant effect.

[17] In the Hurwicz model, $\mu = 0.350$ and $\sigma = 0.050$.

the hypothesis that this new model is consistent with the data cannot be rejected at the 5% level when $a = d = 0.82$ ($S = 0.04$, and $T = 0.11$). Interestingly, the same is true of a model with the sure-thing principle for $\mu = 0.312$, $\sigma = 0.035$, and $a = d = 0.82$. The reason these models have some success is that responses are strongly concentrated in and around the area predicted by the principle of insufficient reason. The Hurwicz criterion with $h = \frac{1}{2}$ is indistinguishable from this principle, and the same goes for the sure-thing principle for subjects who assign RED a probability close to $\frac{1}{3}$.

In summary, our model best fits the data when it describes our population *as if*: (i) each subject is aiming to choose according to the sure-thing principle; (ii) most subjects have an indifference probability of RED close to $\frac{1}{3}$; and (iii) each subject has a moderate tendency to make errors, with errors in an ambiguity-averse direction being somewhat more likely. Indeed, this version of our model cannot confidently be rejected.

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 optimism within the paradigm of the Ellsberg paradox (Pulford [30], Lauriola *et al* [26]). We therefore explored this relationship measuring the degree of optimism with the Life Orientation Test, Revised (LOT-R) of Scheier *et al* [33].

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 behaviour, and a positive result indicating ambiguity-averse behaviour. 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).[18]

We compared these measures of ambiguity aversion and 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* [31]). In contrast to other studies, our measure of optimism was not associated with ambiguity aversion.

Some research suggests that ambiguity aversion differs across genders (Borghans *et al* [5], Powell and Ansic [29]). 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 on the previously mentioned website.

[18] These LOT-R results are consistent with the norms reported in Scheier *et al* [33].

8 Conclusion

Ambiguity aversion has been claimed to be a phenomenon that can be regularly observed in a majority of subjects' choices (Liu and Colman [27], p. 279). We designed a new experiment to examine how much of this phenomenon could be explained by behavior in accordance with the Hurwicz criterion. To our surprise, we found that only the principle of insufficient reason had marked predictive power in our experiment, and that ambiguity aversion was less pronounced than in many other studies. Moreover, we found little evidence of behavior in accordance with the Hurwicz criterion. We twice changed the framing of our experiment, which has 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 criterion for ambiguity aversion is more demanding than is usual in the literature because our criterion requires a subject to display some aversion to the ambiguous option in *two* different but related types of choices. Subjects whose choices do not meet this stricter criterion cannot be said to robustly ambiguity averse.

In summary, our experiment suggests that ambiguity aversion is not always as powerful and robust a phenomenon as it is sometimes said to be.

References

[1] D. Ahn, S. Choi, D. Gale, and S. Kariv. Estimating ambiguity aversion in a portfolio choice experiment. http://www.econ.nyu.edu/user/galed/papers.html, 2010.

[2] S. Becker and F. Browson. What price ambiguity? The role of ambiguity in decision-making. *Journal of Political Economy*, 72:62–73, 1964.

[3] K. Binmore. *Rational Decisions*. Princeton University Press, Princeton, 2009.

[4] J. Bono and T. Judge. Core self-evaluations: A review of the trait and its role in job satisfaction. *European Journal of Personality*, 17:S5–S8, 2003.

[5] L. Borghans and J. Heckman and B. Golsteyn and H. Meijers. Gender differences in risk and ambiguity aversion. *Journal of the European Economic Association*, 7:649–658, 2009.

[6] K. Brewer. Decisions under uncertainty: Comment. *Quarterly Journal of Economics*, 79:657–653, 1965.

[7] C. Camerer and M. Weber. Recent developments in measuring preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, 5:325–370, 1992.

[8] C. Chow and R. Sarin. Comparative ignorance and ambiguity aversion. *Journal of Risk and Uncertainty*, 22:129–139, 2001.

[9] S.P. Curley, J.F. Yates, and A.A. Abrams. Psychological Sources of Ambiguity Avoidance. Organizational Behavior and Human Decision Processes 38, 230–256, 1986.

[10] S. Curley and J. Yates. An empirical evaluation of descriptive models of ambiguity reactions in choice situations. *Journal of Mathematical Psychology*, 33:397–427, 1989.

[11] D. Ellsberg. *The Theory and Practice of Blackmail: Formal Theories of Negotiation*. University of Illinois Press, Urbana, 1975.

[12] D. Ellsberg. *Risk, Ambiguity, and Decision*. Garland Publishing, New York and London, 2002.

[13] J. Etner, M. Jeleva, and J-M. Tallon. Decision theory under ambiguity. *Journal of Economic Surveys*, 24:1–45, 2010.

[14] W. Fellner. Distortion of subjective probabilities as a reaction to uncertainty. *American Economic Review*, 48:670–689, 1961.

[15] C. Fox and A. Tversky. Ambiguity aversion and comparative ignorance. *Quarterly Journal Economics*, 110:585–603, 1995.

[16] D. Frisch and J. Baron. Ambiguity and rationality. *Journal of Behavioral Decision Making*, 1:149–157, 1988.

[17] P. Ghirardato and M. Marinacci. Ambiguity made precise: A comparative foundation. *Journal of Economic Theory*, 102:251–289, 2002.

[18] I. Gilboa. *Uncertainty in Economic Theory: Essays in Honor of David Schmeidler's 65th Birthday*. Routledge, London, 2004.

[19] Y. Halevy. Ellsberg revisited: An experimental study. *Econometrica*, 75:503–506, 2007.

[20] J. Hey, G. Lotito, and A. Maffioletti. The descriptive and predictive adequacy of decision making under uncertainty/ambiguity. *Journal of Risk and Uncertainty*, 41:81–111, 2010.

[21] M. Hsu, M. Bhatt, R. Adolphs, D. Tranel, and C. Camerer. Neural systems responding to degrees of uncertainty in human decision-making. *Science*, 310:1680–1883, 2005.

[22] L. Hurwicz. Optimality criteria for decision making under ignorance. Cowles Commission Discussion Paper, Statistics 370, 1951.

[23] J. Kadane. Healthy scepticism as an expected-utility explanation of the phenomena of Allais and Ellsberg. In *Decision-Making under Risk and Uncertainty: New Models and Empirical Findings*. Kluwer, Dordrecht, 1992. (Edited by J. Geweke).

[24] D. Kahneman and A. Tversky. The framing of decisions and the psychology of choice. *Science*, 211:453–458, 1981.

[25] G. Keren and L. Gerritsen. On the robustness and possible accounts of ambiguity aversion. *Acta Psychologica*, 103:149–172, 1999.

[26] M. Lauriola, I.Levin, and S. Hart. 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, 2007.

[27] H-H. Lui and A. Colman. Ambiguity aversion in the long run: Repeated decisions under risk and uncertainty. *Journal of Economic Psychology*, 30:277–284, 2009.

[28] J. Milnor. Games against Nature. In *Decision Processes*. Wiley, New York, 1954. (Edited by R. Thrall, C. Coombs, and R. Davies).

[29] M. Powell and D. Ansic. Gender differences in risk behavior in financial decision-making. *Journal of Economic Psychology*, 18:605–828, 1997.

[30] B. Pulford. Is luck on my side? Optimism, pessimism, and ambiguity aversion. *Quarterly Journal of Experimental Psychology*, 62:1079–1087, 2009.

[31] J. Rouder, P. Speckman, D. Sun, R. Morley, and G. Iverson. Bayesion t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, 62:1079–1087, 2009.

[32] L. Savage. *The Foundations of Statistics*. Wiley, New York, 1954.

[33] M. Scheier, F. Carver, and M. Bridges. 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, 1994.

[34] H. Schneeweiss. Spieltheoretische analyse des Ellsberg-paradaxons. *Zeitschrift fur die Gesamte Staatswissenschaft*, 124:249–255, 1968.

[35] A. Sproten, C. Diener, C. Fiebach, and C. Schwieren. Ageing and decision making: How ageing affects decisions under uncertainty. Working Paper 508, Economics Department, University of Heidelberg, 2010.

[36] T. Trautman, F. Vieider, and P. Wakker. Causes of ambiguity aversion: Known versus unknown preferences. *Journal of Risk and Uncertainty*, 36:225–243, 2008.

[37] A. Wald. *Statistical Decision Theory*. Wiley, New York, 1950.