

Original citation:

Li, Zhihua, Loomes, Graham and Pogrebna, Ganna. (2017) Attitudes to uncertainty in a strategic setting. *Economic Journal*, 127 (601). pp. 809-826.

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/84374>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution 4.0 International license (CC BY 4.0) and may be reused according to the conditions of the license. For more details see: <http://creativecommons.org/licenses/by/4.0/>

A note on versions:

The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

ATTITUDES TO UNCERTAINTY IN A STRATEGIC SETTING*

Zhihua Li, Graham Loomes and Ganna Pogrebna

Much uncertainty in life relates to the behaviour of others in interactive environments. This article tests some implications of subjective expected utility theory (Savage, 1954) in an experimental strategic setting where there is uncertainty about the actions of other players. In this environment, a large majority of our participants violate subjective expected utility theory. However, they do not exhibit the sorts of consistent ‘attitude to ambiguity’ found in individual decision experiments. We discuss three possible explanations of their behaviour: non-linear transformation of probabilities; noise in responses; and/or systematic biases in the way that individuals generate subjective probabilities.

Much of the evidence about failures of Savage’s (1954) subjective expected utility theory (SEUT) – and in particular, much of the support for the ‘ambiguity aversion’ reported by Ellsberg (1961) – comes from experiments involving individual decisions in ‘games against nature’.¹ The bulk of these data relate to non-interactive environments where ambiguity is implemented as a two-stage lottery (Bernasconi and Loomes, 1992; Maafi, 2011); or as a lottery with unknown probability of an ambiguous event while the joint probability of this event with one other event is known (Halevy, 2007); or as a lottery where a rough idea about the probability of an event can be inferred from an experiential task (Hey *et al.*, 2010); or as betting upon some natural phenomenon such as the temperature in a specified location at a particular time (Baillon, 2008).

However, in many important economic, political and social situations, uncertainty arises from the difficulty of predicting the actions of other agents rather than the composition of coloured balls in a bag or in a bingo blower. It has been suggested that the source of uncertainty may affect the nature and degree of any attitude to ambiguity (Abdellaoui *et al.*, 2011) and there may be reasons why individuals making choices involving lotteries devised by an experimenter might behave differently than when they

* Corresponding author: Graham Loomes, Warwick Business School, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK. Email: Graham.Loomes@wbs.ac.uk.

We are grateful to the attendees at the Decision Research at Warwick (DR@W) forum, the Foundations of Utility and Risk Conference in Rotterdam (2014), the Decision Making Conference in Bristol (2014), the Network for Integrated Behavioural Science (NIBS) in Nottingham (2015) and the Royal Economic Society Conference at the University of Sussex (2016) for many insightful comments and suggestions. We also appreciate the helpful suggestions made in the course of the editorial process. We acknowledge support from ESRC Award ES/K002201/1 and from the Leverhulme Trust ‘Value’ Programme RP2012-V-022. Ganna Pogrebna acknowledges financial support from the RCUK/EPSRC project EP/N028422/1 and RCUK/ EPSRC project EP/L023911/1.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

¹ In his book, *The Foundations of Statistics*, Savage integrated a normative account of decision making under risk (when probabilities are known) and under uncertainty (when probabilities are not objectively known). In doing so, his SEUT postulated that individuals have neutral attitudes towards uncertainty, so that beliefs about the probabilities of uncertain events could be extracted from observed choices (Machina and Schmeidler, 1992). However, Ellsberg (1961) challenged the validity of SEUT by providing examples where individuals exhibited non-neutral attitudes towards uncertainty.

are forming beliefs about the behaviour of other individuals. For example, there has long been a concern that (at least some) participants may believe that the experimenter is trying to manipulate their behaviour and/or minimise their payoffs (see the discussion about ‘allaying suspicion’ in Binmore *et al.*, 2012, pp. 223–4). Or if people are processing the tasks as compound lotteries, it may be that a failure of the reduction of compound lotteries axiom is being interpreted as aversion to uncertainty *per se*.² Or it may be that some participants have limited numeracy skills and/or lack motivation to think about more complex options, and this may be expressed by choosing a transparent lottery with known risks rather than a more opaque prospect that may require more careful thought.

In order to investigate whether the patterns found in classic non-interactive scenarios carry over into an interactive environment, we devised a rather different experimental format. The essence of our design involved asking participants to play a co-ordination game and then eliciting their subjective probabilities concerning the relative likelihoods of the strategies chosen by the other participants in the same session. Hence the source of the uncertainty was independent of the experimenters and immune from manipulation by them. Moreover, there was nothing in our design that prompted compound-*versus*-single-stage lottery comparisons which might be confounded with attitudes to uncertainty. Rather, we aimed to create an environment where the uncertainty emanated entirely from the actions of other people facing the same task but conceivably arriving at somewhat different decisions.

A number of previous experiments have elicited beliefs in games of various kinds, but they have typically imposed the SEUT requirement that the judged probabilities of the set of mutually exclusive and collectively exhaustive strategies should sum to 1. When we do not impose that restriction, we find evidence of highly systematic departures from SEUT. However, these departures look rather different from the aversion to ambiguity exhibited in many non-interactive individual choice experiments. The patterns we find are more complex, somewhat contextual, and may be better understood in terms of (a variant of) Tversky and Koehler’s (1994) Support Theory.

The remainder of this article is organised as follows. The basic idea behind the experiment and the key SEUT-based hypotheses are set out in Section 1. We describe the experimental design and its implementation in more detail in Section 2. Analysis of the data is provided in Section 3. In Section 4, we discuss several possible accounts of the data, together with some supplementary analysis. Finally, we draw some conclusions in Section 5.

1. Hypotheses

The experiment (described in more detail in the next Section) revolves around a form of co-ordination game where players are presented with several sets of three items and,

² Halevy (2007, p. 532) concludes that ‘failure to reduce compound (objective) lotteries is the underlying factor of the Ellsberg paradox’. Abdellaoui *et al.* (2015) report a positive, but weaker, association between attitude to compound lotteries and attitude to ambiguity, suggesting that even in individual decision scenarios of the classic kind, there is some room for separating the two.

for each set, are asked to rank the three items in the way they judge most likely to match the ranking of another player chosen at random from among the other participants in the same experimental session. After all players have completed this initial task, there follow a further 15 questions which focus exclusively upon the chances that a randomly selected other player would have put a particular item at the top of his/her ranking (equal rankings were not allowed).

If we denote the three items in any set by X , Y and Z , we can define three possible states of the world: S_X , where the randomly selected other participant put X top of their ranking; S_Y , where the other person placed Y first in their ranking; and S_Z , where the other person put Z top. An individual's beliefs about the likelihood of each state occurring are denoted, respectively, by $\pi(X)$, $\pi(Y)$ and $\pi(Z)$. Schelling (1980) suggested, and a number of subsequent experiments have confirmed (Mehta *et al.*, 1994; Bardsley *et al.*, 2010), that the natures and/or descriptions of X , Y and Z might result in people judging some $\pi(\cdot)$ s to be higher or lower than others. SEUT is silent on that and simply entails that $0 \leq \pi(X)$, $\pi(Y)$, $\pi(Z) \leq 1$ and that $\pi(X) + \pi(Y) + \pi(Z) = 1$.

For each set, we construct three state-contingent claims $C(X)$, $C(Y)$ and $C(Z)$, each of which pays £20 if the other person turns out to have put the item in question at the top of their ranking but pays 0 otherwise. We can also construct lotteries involving a well-specified random mechanism generating some 'objective' probability p of receiving £20 and a $1 - p$ probability of 0. Then for each of $C(X)$, $C(Y)$ and $C(Z)$, there will be a 'probability equivalent' lottery $PE(X)$, $PE(Y)$ and $PE(Z)$ such that the individual is indifferent between $C(i)$ and $PE(i)$ for $i = X, Y, Z$. The probabilities of £20 which produce these equivalences are denoted by $p(X)$, $p(Y)$ and $p(Z)$.

We also construct state-contingent claims on the unions of any two states of the world, e.g. the claim $C(X \cup Y)$ that pays £20 if the other person turns out to have put either X or Y top but pays 0 if he put Z top; and likewise for $C(X \cup Z)$ and $C(Y \cup Z)$. For these three claims, the probability equivalent lotteries are $PE(X \cup Y)$, $PE(X \cup Z)$ and $PE(Y \cup Z)$, with the indifference values denoted by $p(X \cup Y)$, $p(X \cup Z)$ and $p(Y \cup Z)$.

Under SEUT assumptions of ambiguity neutrality, we expect $\pi(i)$ to be revealed by $p(i)$, so that we have the following hypotheses:

$$H1: p(X) + p(Y) + p(Z) = 1.$$

$$H2: p(X \cup Y) = p(X) + p(Y) \text{ for all } \{X, Y\}; \text{ which, in conjunction with H1, implies:}$$

$$H3: p(X \cup Y) + p(X \cup Z) + p(Y \cup Z) = 2.$$

However, if – as is often supposed – those who are not ambiguity neutral tend to be ambiguity averse, we should expect a preference for known probabilities over uncertain beliefs to produce $p(X) + p(Y) + p(Z) < 1$ and also $p(X \cup Y) + p(X \cup Z) + p(Y \cup Z) < 2$. Ambiguity seeking would produce inequalities in the opposite direction. It is not obvious that ambiguity attitude *per se* provides a particular alternative hypothesis to H2 but the data relating to this hypothesis may nevertheless be of interest.

2. The Experiment

2.1. *Experimental Design*³

We constructed eight sets of three items using the following categories: METALS, COLOURS, FRUITS, ANIMALS, FLOWERS, PETS, GEMS, TRANSPORT. To keep experimental sessions to less than 1 hour, we divided participants into two treatment subsamples, T1 and T2. All participants went through the same Practice phase where we used a set (METALS) which was common to both subsamples. After the Practice phase, all subsequent decisions were incentivised in ways explained below. One set of three items – COLOURS – was common to both subsamples. Members of T1 also saw sets involving FLOWERS, FRUITS, and ANIMALS, while members of T2 saw GEMS, TRANSPORT and PETS. Table 1 summarises this information and lists the specific items within each set.

For each set in turn, each participant was presented with a series of 16 questions through the medium of a computer screen, with each individual in a separate laboratory cubicle so that no communication was possible. Although the order in which items were displayed on the screen was randomised, the sequence of 16 questions was always in the same order. The questions which are central to the tests of hypotheses H1, H2 and H3 are Questions 11–16 in conjunction with the session responses to Question 1. Questions 2–10 in each series were intended to give participants plenty of opportunity to consider and possibly refine their judgments before they gave their Questions' 11–16 responses.

In order to give readers a clear and complete picture of what participants saw, we now briefly outline all types of questions in the series and indicate their purposes.

Question 1 asked participants to rank the three items from the one they thought most likely to be ranked top by a randomly chosen other player who was trying to co-ordinate on the same ranking, down to the one they ranked third. Even though our focus in all subsequent questions was to be upon the top-ranked item, we wanted

Table 1
Sets Used in the Experiment

| Set | Items | Subsample |
|-----------|---------------------------|--------------------|
| METALS | Copper, Gold, Iron | Practice T1 and T2 |
| COLOURS | Blue, Green, Red | T1 and T2 |
| FLOWERS | Daisy, Rose, Tulip | T1 |
| FRUITS | Apple, Banana, Strawberry | T1 |
| ANIMALS | Giraffe, Hippo, Lion | T1 |
| GEMS | Diamond, Emerald, Ruby | T2 |
| TRANSPORT | Bicycle, Bus, Car | T2 |
| PETS | Cat, Dog, Rabbit | T2 |

Notes. Table 1 shows all sets and items used in the two subsamples T1 and T2. Both subsamples answered practice questions using the METALS set. All other questions were incentivised, with COLOURS common to both subsamples.

³ Detailed experimental instructions and screenshots of each type of question are provided in online Appendix A.

participants to think carefully about the salience of all three items in each set and the incentive mechanism was intended to achieve that. So if a question of this type was selected as a basis for paying a particular individual, a position (first, second, third) was chosen at random by rolling a three-sided die. Then the item placed in that position by this individual was compared with the item placed in the same position by one other randomly chosen participant. If it was the same item, this individual received £20; otherwise he received nothing.

The responses to this question collectively determine the probabilities of the various states of the world in the subsequent questions. To illustrate, suppose that from a particular individual's perspective there are 20 participants in the same session and that, when answering Question 1 for the set METALS, 13 of those other participants put Gold first in their ranking, 5 put Iron top and 2 place Copper first. Then from this individual's perspective, the probabilities that a randomly chosen other player will have put Gold or Iron or Copper first in their ranking are, respectively, 0.65, 0.25 and 0.10. Of course, the individual does not know that these are the probabilities but rather must operate on the basis of his/her beliefs about the distribution of those top rankings. It is these beliefs that subsequent questions aim to tap into, culminating in attempts to elicit them directly and precisely in Questions 11–16.

Question 2 in each series provided individuals with a 'budget' of £20 and asked them to distribute this sum between the three items, bearing in mind how likely they thought it was that each item would have been ranked top in Question 1 by another randomly chosen participant. If this question were picked to be the basis for payment, the first-ranked item of another randomly selected participant would be revealed and the individual would receive whatever amount he/she had allocated to that item. So this task encouraged individuals to think about their subjective probabilities, since under most circumstances the amounts allocated to different items will be positively correlated with their respective π s.⁴

Questions 3–8 introduced the idea of claims contingent upon each state of the world. The display reproduced in online Appendix A shows the case where the individual is given a claim that will pay £20 if a randomly chosen other participant turns out to have put Lion at the top of his/her ranking but will pay 0 if that participant has put either Giraffe or Hippo top. The question then elicits the individual's certainty equivalent (CE) for that claim: that is, the sure sum of money that the individual would regard as exactly as desirable as playing out the claim and getting paid accordingly. Questions 3, 4 and 5 in each series elicited CEs for claims contingent upon each of the single items in the set. Questions 6, 7 and 8 elicited CEs for claims contingent on each of the possible pairings of items. A standard Becker–DeGroot–Marschak (BDM) incentive mechanism was used (see Becker *et al.*, 1964 and the example in online Appendix A for details of the wording and procedure). These six questions gave individuals further incentives to think about and possibly refine their judgments about the likelihoods of the different events.

⁴ This type of 'allocation task' has been used in earlier studies involving well-specified random mechanisms – see Loomes and Pogrebná (2014) for a recent example and discussion – and allocations appear to be reasonably sensitive to variations in probabilities, although there are 'rounding' effects which tend to limit the precision of this instrument.

Question 9 then asked each respondent for a direct estimate of the distribution of top rankings among ten other randomly selected participants in the same session. Each respondent was told that if this question were selected as the basis for payment, one item would be selected at random and his or her estimate would be compared with how many out of those other ten participants put that item first in Question 1: if the estimate was correct, the respondent was paid £20; if the estimate was incorrect, they received 0. This provides a very strong incentive to think carefully about the distribution of other players' top rankings. However, it can only produce rather coarse estimates that would translate to probabilities expressed in multiples of 0.1, so it was followed by a question that gave respondents an opportunity to give more fine-grained answers.

Question 10 asked respondents to distribute 100% across the three items according to their belief about the likelihood that one other randomly chosen participant would select each of the items as their top choice in Question 1. If this question were selected to be the basis of payment, one of the items would be selected at random and a lottery would be constructed offering the same objective probability of £20 as the probability the participant had allocated to that item. Then a coin would be spun to determine whether he was paid according to the lottery or by seeing whether another randomly selected participant had top-ranked that item.

Within the terms of our design, it was not possible to make it a strictly dominant strategy to state one's best judgment about the probabilities to the nearest 1%.⁵ But the primary purpose of Question 10 was to encourage respondents to think carefully about the probabilities ahead of the crucial PE questions and refine their answers to Question 9. In general, that objective was achieved, as we shall see in due course. However, even if some individuals failed to report their beliefs accurately in Question 10, it would not adversely affect the tests of hypotheses H1–H3, which do not use these data.⁶ Those tests depend solely on the data from the final six questions in each series which we now describe.

Questions 11–16 involved the same six state contingent claims that had featured in questions 3–8, except that now it was PEs rather than CEs for those claims that were being elicited. Questions 11, 12 and 13 presented claims contingent upon each of the three single items and for each in turn asked participants to indicate via a slider the best objective lottery that they would (just) reject in order to play out the claim and the smallest objective chance of £20 (1% higher) that they would (just) prefer to play instead of the claim. A standard BDM mechanism was used to provide the incentive to give an accurate and truthful answer. The same procedure was used in questions 14, 15 and 16, where the claims were contingent upon each possible pair of items from that

⁵ To achieve incentive compatibility along the lines of Question 9 would have required us to have had the responses of at least 100 other players available from the first session onwards and that was not practicable.

⁶ We have been asked whether the Question 10 task might have 'led' respondents to think that their subjective probabilities, as elicited in the subsequent three questions in each series, 'ought' to sum to 1. This might or might not have been the case. However, if it were the case, it would not be the source of departures from standard assumptions: on the contrary, it would be giving SEUT's implications, and in particular H1, every chance to succeed. If the PEs elicited subsequently exhibit systematic departures from SEUT, it will be harder to attribute those departures to lack of opportunity to think about making the PEs sum to 1 if that is what respondents want to do.

set. It is the data from these final six questions for any set of items that provide the basis for our tests of H1–H3.

2.2. *Experimental Implementation*

We recruited 105 participants from the electronic database of the Decision Research at Warwick (DR@W) subject pool. All participants were undergraduate students at the University of Warwick. Of these, 51 were randomly assigned to T1 and 54 to T2. The experiment was computerised, using the EXPERT software.

At the beginning of the experiment, each participant drew a unique ID which was used by the computer program to randomly match participants with each other at the end of the experiment. They received no other information about their counterpart(s).

After responding to 64 decision problems (four series of 16) and completing a short questionnaire, each participant drew a number from an opaque bag containing numbers from 1 to 64. The computer displayed the decision problem corresponding to the randomly drawn number and the participant could see that question along with his or her answer. The question was then played out according to the instructions shown on the computer screen. Each experimental session lasted approximately 1 hour: on average, each participant received about £11.

3. Results

Table 2 provides basic statistics – medians and means – for the various types of questions applied to single items. There is a high degree of concordance across the different question types. In some sets, one item is clearly a strong favourite while the other two are about equally – but relatively poorly – favoured; and this relationship comes through in all types of questions – see FLOWERS and GEMS for examples. In other sets, two items vie for first choice, with the third item some way behind – see FRUITS and COLOURS – and this pattern is also broadly consistent across all question types.

We now turn to the main focus of the article: the extent to which the beliefs expressed via probability equivalents are consistent with SEUT. We start with the first and third hypotheses specified in Section 1, namely:

$$\text{H1: } p(X) + p(Y) + p(Z) = 1;$$

$$\text{H3: } p(X \cup Y) + p(X \cup Z) + p(Y \cup Z) = 2.$$

To examine H1, for each individual we sum the responses to questions 11, 12 and 13 within each category set. To examine H3, we sum the responses to questions 14, 15 and 16. Since the questions were presented to participants in terms of percentages rather than probabilities, we shall from now on express probabilities in percentage form.

The theory and the hypotheses are formulated on the basis of deterministic, noise-free preferences and beliefs. In reality, however, most people's responses to experimental tasks are subject to some degree of noise and/or imprecision. If such variability were symmetrical around core preferences that conform with SEUT, we

Table 2
Basic Statistics

| Sub-sample | Set | Items | Median (average) value obtained per single item | | | | | |
|------------|-----------|------------|-------------------------------------------------|------------------|------------------|-------------|-----------------|-----------------|
| | | | Q1 | Q2 | Q3–5 | Q9 | Q10 | Q11–13 |
| | | | Ranked top % | Alloc £ | CE £ | 10 others | Alloc % | PE % |
| T1 | FLOWERS | Rose | 88.2 | 12.30 (12.33) | 12.51 (12.18) | 7 (6.59) | 60 (59.63) | 74 (68.98) |
| | | Tulip | 5.9 | 3.00 (3.51) | 4.09 (4.51) | 2 (1.53) | 20 (18.76) | 25 (28.45) |
| | | Daisy | 5.9 | 3.00 (4.16) | 3.67 (4.33) | 2 (1.88) | 19 (21.61) | 20 (27.76) |
| T1 | ANIMALS | Lion | 92.2 | 11.70 (12.03) | 11.34 (12.06) | 6 (6.49) | 65 (61.78) | 70 (66.69) |
| | | Giraffe | 5.9 | 4.60 (4.71) | 5.91 (5.77) | 2 (2.18) | 23 (22.55) | 34 (34.27) |
| | | Hippo | 2.0 | 3.00 (3.26) | 4.42 (5.04) | 1 (1.33) | 11 (15.67) | 27 (28.63) |
| T1 | FRUITS | Apple | 45.1 | 6.70 (7.17) | 7.97 (8.76) | 3 (3.71) | 31 (35.94) | 45 (49.06) |
| | | Strawberry | 47.1 | 7.80 (8.15) | 8.51 (9.11) | 5 (4.37) | 38 (42.22) | 48 (48.86) |
| | | Banana | 7.8 | 5.00 (4.68) | 5.55 (5.93) | 2 (1.92) | 24 (21.84) | 34 (35.27) |
| T1 | COLOURS | Red | 70.6 | 8.00 (8.66) | 8.90 (9.41) | 5 (4.65) | 40 (42.10) | 51 (54.35) |
| | | Blue | 29.4 | 6.60 (6.94) | 6.99 (8.37) | 3 (3.63) | 34 (37.31) | 47 (48.59) |
| | | Green | 0 | 5.00 (4.40) | 5.13 (5.96) | 2 (1.72) | 20 (20.59) | 28 (29.59) |
| T2 | GEMS | Diamond | 94.4 | 15.00 (14.24) | 15.25 (13.24) | 8 (7.65) | 79.5 (74.06) | 79.5 (70.33) |
| | | Ruby | 5.6 | 2.45 (2.99) | 4.99 (5.20) | 1 (1.30) | 10 (13.22) | 30 (33.02) |
| | | Emerald | 0 | 2.40 (2.76) | 3.64 (4.34) | 1 (1.06) | 10 (12.72) | 20 (23.96) |
| T2 | TRANSPORT | Car | 79.6 | 12.40 (12.31) | 14.51 (12.85) | 7 (6.54) | 65.5 (61.91) | 64 (60.22) |
| | | Bus | 16.7 | 4.00 (3.84) | 5.00 (5.63) | 1 (1.72) | 13.5 (18.43) | 25.5 (30.54) |
| | | Bicycle | 3.7 | 4.00 (3.85) | 4.95 (5.68) | 1 (1.74) | 19.5 (19.67) | 25 (29.04) |
| T2 | PETS | Dog | 72.2 | 9.90 (10.19) | 10.11 (10.71) | 5 (5.46) | 50 (53.04) | 59 (60.04) |
| | | Cat | 20.4 | 6.70 (6.30) | 6.86 (7.05) | 3 (3.33) | 32.5 (32.63) | 36.5 (36.61) |
| | | Rabbit | 7.4 | 2.90 (3.51) | 3.48 (4.49) | 1 (1.20) | 10 (14.33) | 17 (21.48) |
| T2 | COLOURS | Red | 40.7 | 6.95 (7.86) | 8.05 (7.73) | 4 (4.22) | 34 (40.52) | 40 (44.07) |
| | | Blue | 51.9 | 6.80 (7.30) | 8.05 (8.05) | 4 (3.72) | 34 (36.80) | 43.5 (42.35) |
| | | Green | 7.4 | 5.00 (4.83) | 5.52 (5.68) | 2 (2.06) | 23.5 (22.69) | 29 (28.19) |

Notes. Table 2 provides basic statistics obtained from those questions (Q1–Q5 and Q9–Q13) which relate to single items (Q1–Q13 refer to Questions 1–13 respectively). Ranked Top %, percentage of respondents who put the item first in their ranking in Q1; Alloc £, amount of money out of £20 placed on an item in Q2; CE £, certainty equivalent for each item from Q3 to Q5; 10 others, belief about the number of other players (out of 10 randomly selected) placing item top in their Q1 response; Alloc %, belief about the probability of each item being placed top by a randomly selected other player when the sum of the probabilities per set is constrained to sum to 100%; PE %, probability equivalent for each item being placed top in the ranking when the probabilities per set were not constrained to sum to 100%.

Table 3

Sum of Probability Equivalents of Single-item Claims Versus Sum of Probability Equivalents for Pair-based Claims Within Each Set

| Sub-sample | Set | Summary | | | | Two-tailed t-tests of mean differences |
|------------|-----------|-----------------------------|-------------------------------|------|-------|----------------------------------------------------------------------------------------------------------------------|
| | | Sum of PEs for single items | Sum of PEs for pairs of items | | | |
| | | | ≤200 | >200 | Total | |
| T1 | FLOWERS | ≤100 | 12 | 0 | 12 | Sum of PEs for single items Mean: 125.20 (p < 0.001) Sum of PEs for pairs of items Mean: 186.92 (p = 0.034) |
| | | >100 | 20 | 19 | 39 | |
| | | Total | 32 | 19 | 51 | |
| T1 | ANIMALS | ≤100 | 8 | 2 | 10 | Sum of PEs for single items Mean: 129.59 (p < 0.001) Sum of PEs for pairs of items Mean: 186.75 (p < 0.01) |
| | | >100 | 25 | 16 | 41 | |
| | | Total | 33 | 18 | 51 | |
| T1 | FRUITS | ≤100 | 9 | 2 | 11 | Sum of PEs for single items Mean: 133.20 (p < 0.001) Sum of PEs for pairs of items Mean: 185.33 (p = 0.052) |
| | | >100 | 22 | 18 | 40 | |
| | | Total | 31 | 20 | 51 | |
| T1 | COLOURS | ≤100 | 12 | 0 | 12 | Sum of PEs for single items Mean: 132.53 (p < 0.001) Sum of PEs for pairs of items Mean: 185.78 (p = 0.016) |
| | | >100 | 24 | 15 | 39 | |
| | | Total | 36 | 15 | 51 | |
| T2 | GEMS | ≤100 | 11 | 5 | 16 | Sum of PEs for single items Mean: 127.31 (p < 0.001) Sum of PEs for pairs of items Mean: 182.37 (p = 0.013) |
| | | >100 | 19 | 19 | 38 | |
| | | Total | 30 | 24 | 54 | |
| T2 | TRANSPORT | ≤100 | 17 | 0 | 17 | Sum of PEs for single items Mean: 119.80 (p < 0.001) Sum of PEs for pairs of items Mean: 179.07 (p < 0.001) |
| | | >100 | 22 | 15 | 37 | |
| | | Total | 39 | 15 | 54 | |
| T2 | PETS | ≤100 | 15 | 3 | 18 | Sum of PEs for single items Mean: 118.13 (p < 0.001) Sum of PEs for pairs of items Mean: 177.70 (p < 0.001) |
| | | >100 | 22 | 14 | 36 | |
| | | Total | 36 | 17 | 54 | |
| T2 | COLOURS | ≤100 | 20 | 2 | 22 | Sum of PEs for single items Mean: 114.61 (p < 0.01) Sum of PEs for pairs of items Mean: 168.50 (p < 0.001) |
| | | >100 | 23 | 9 | 32 | |
| | | Total | 43 | 11 | 54 | |

Notes. For each set, cross-tabulations show the number of people whose sum of probability equivalents (PEs) for three single items as elicited in Questions 11–13 is greater than 100% or less than or equal to 100% versus those whose sum of PEs for 3 pairs of items as elicited in Questions 14–16 is greater than 200% or less than or equal to 200%. Two-tailed t-tests of whether single-item sums differ from 100% and whether paired-item sums differ from 200% are reported in the last column.

should expect, on average, as many cases where individuals' $p(X) + p(Y) + p(Z)$ add up to more than 100% as cases where they sum to less than 100% (with a few cases, perhaps, where they happen to add up to exactly 100%). Likewise, we should expect a few cases where $p(X \cup Y) + p(X \cup Z) + p(Y \cup Z) = 200\%$, while the majority of cases should be liable to fall in roughly equal numbers on either side of 200%.

As it turned out, there were 22 cases out of 420 (105 participants × 4 category sets) where $p(X) + p(Y) + p(Z)$ summed to exactly 100%; and there were 16 cases out of 420 where $p(X \cup Y) + p(X \cup Z) + p(Y \cup Z)$ summed to exactly 200%. Given the low

frequencies involved, we combine such cases with those where the sums are less than 100% or less than 200% to produce the summary cross-tabulations in Table 3, which sort individuals according to the conjunction of their PE sums for single items and their PE sums for pairs of items. The patterns are quite robust across all eight cross-tabulations.

Typically, slightly more than half of the observations fall into either the top-left or bottom-right cells, seeming either to underweight or else overweight both the single events and their unions. However, just under half of the respondents fall into the off-diagonal cells and, in every case, do so in a highly asymmetrical manner, such that the great majority have single-item sums higher than 100% in conjunction with paired-item sums lower than 200%.

Aggregating across individuals within categories, the right-hand column of Table 3 reports the observed subsample means for single-item sums and paired-item sums, together with two-tailed t-tests of the hypotheses for each category.

We see that the single-item means are all greater than 100%, with the null being rejected at the 0.01 or 0.001 level in every case; at the same time, the paired-item means are all less than 200% and reject the null at the 0.05 level, except for FRUITS, where $p = 0.052$.

Had we been observing only the paired-item data, it might have seemed that individuals act as if predominantly ambiguity averse, with a strong tendency to accept lotteries with known risks lower than the probabilities of the state-contingent claims inferred from the allocation task. However, the data from the single-item PEs show an asymmetry that is at least as strong in the opposite direction: that is, in the direction which would normally be associated with ambiguity seeking.

Averaging individual responses across sets produces a pattern which is, if anything, even sharper. Table 4 sorts all 105 participants according to their individual mean sum of PEs for single-item claims and their mean sum of PEs for paired-item claims.

If we were to regard the 15 individuals in the $\{<100, <200\}$ cell as ambiguity averse and the 31 in the $\{>100, >200\}$ cell as ambiguity seeking, we would conclude – contrary to much of the previous literature – that ambiguity seeking is more common than ambiguity aversion. However, these two cells combined are outweighed by the 57 who

Table 4

Individuals' Mean Sum of Probability Equivalents for Single-item Claims Versus Mean Sum for Pair-based Claims

| | | Individuals' mean sums of PEs for paired-item claims | | | |
|------------------------------------------------------|-------|------------------------------------------------------|------|------|-------|
| | | <200 | =200 | >200 | Total |
| Individuals' mean sums of PEs for single-item claims | <100 | 15 | 0 | 1 | 16 |
| | =100 | 1 | 0 | 0 | 1 |
| | >100 | 57 | 0 | 31 | 88 |
| | Total | 73 | 0 | 32 | 105 |

Notes. Table 4 provides a summary of the individual mean sums of PEs for single and paired items. The population of 105 individuals is mapped according to individual responses in Questions 11–13 *versus* Questions 14–16.

manifest the predominant $\{>100, <200\}$ ‘mix’ of inequalities (with only 1 individual mixed in the opposite way).

So both H1 and H3 are clearly rejected but in a manner that suggests a tendency among a substantial number of individuals to treat single-item claims rather differently than their paired-item counterparts. Moreover, even within the 15 and 31, there are quite a few for whom it is very often the case that $p(X) + p(Y) > p(X \cup Y)$. These data appear to be inconsistent with the other hypothesis specified earlier, namely:

$$\text{H2: } p(X \cup Y) = p(X) + p(Y) \text{ for all } \{X, Y\}$$

To test H2 directly, for each $\{X, Y\}$ pair we compute each individual's $[p(X) + p(Y) - p(X \cup Y)]$. If the only deviations from the equalities in H2 are due to white noise, we should expect sample means for each pair to be insignificantly different from zero. On the basis of the data in Tables 3 and 4, however, we should expect those sample means to be significantly positive for each $\{X, Y\}$ pair.

Table A1 in online Appendix A reports the results in detail. For every $\{X, Y\}$ pair, the great majority of observations – typically around 80% of individuals – involve the sum of $p(X)$ and $p(Y)$ being greater than $p(X \cup Y)$. The smallest subsample mean difference – for {Tulip, Daisy} – is 16.78 percentage points, while the largest mean difference – for {Red, Blue} in T1 – is 31.10 percentage points. Tests of the hypothesis $[p(X) + p(Y) - p(X \cup Y)] = 0$ reject the null so strongly that $p < 0.001$ in every case; hence Table A1 reports instead the t-statistics, with the smallest of these – for {Car, Bus} – being 4.61.

When analysis is conducted at the level of the individual, it shows that the phenomenon is pervasive. For each individual we have 12 $\{X, Y\}$ pairs and can compute the individual's mean and median values of $[p(X) + p(Y) - p(X \cup Y)]$ for those 12 pairs together. In T1 there is just one individual for whom the mean difference across the 12 pairs is ≤ 0 , and six others for whom the median is ≤ 0 . In T2 there are five individuals for whom the mean and/or the median is ≤ 0 . There are 79 individuals (that is, just over three-quarters of the whole sample) for whom the mean and median values of $[p(X) + p(Y)]$ exceed the corresponding $p(X \cup Y)$ by at least 5 percentage points. In short, H2 is strongly and consistently rejected: for all sets and for the great majority of individuals, $p(X) + p(Y) > p(X \cup Y)$.

In the next Section, we consider three possible ways of accounting for the data reported above.

4. Possible Explanations

There would appear to be several rather different – but not necessarily mutually exclusive – ways of interpreting the observed patterns. One possible account involves systematic transformation of probabilities of the kind proposed by different forms of Prospect Theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) or other rank-dependent models (Quiggin, 1982), where smaller probabilities are overweighted and larger probabilities are underweighted. Another possibility is that uncertainty leads to asymmetric noise in probability judgments and that this results in asymmetric departures from SEUT. A third possibility is that probability judgments are made in a manner consistent with the kinds of psychological insights offered by Support Theory (Tversky

and Koehler, 1994; Rottenstreich and Tversky, 1997; Brenner *et al.*, 2002). Because these accounts have potentially overlapping implications, we cannot provide a sharp discriminatory test between them. However, our results may still give some broad indication of the relative merits of the three candidate explanations.

4.1. Probability Transformation

The notion that decision makers may act as if transforming probabilities non-linearly into decision weights has a long history. It is often supposed that for many individuals and for the ‘representative agent’, such transformations are best modelled by functions taking an inverse-S shape. Gonzalez and Wu (1999) provided a useful overview of a variety of possible functional forms of this type. They were inclined to favour two-parameter functions such as those proposed by Goldstein and Einhorn (1987) and by Prelec (1998), where one parameter relates to elevation and the other determines curvature. Briefly, elevation reflects what Gonzalez and Wu (1999, p. 138) interpret as attractiveness and what Abdellaoui *et al.* (2011, p. 701) regard as optimism/pessimism, while curvature – especially with respect to the slope across the middle range of probabilities – reflects what Gonzalez and Wu (1999, pp. 136–8) describe as discriminability and what Abdellaoui *et al.* (2011, pp. 701–2) call likelihood insensitivity. Both papers report the inverse-S shape to be widely applicable to their own data; and, although there was a good deal of interpersonal variability in terms of both slope and elevation, those data broadly supported the results from earlier studies that the typical pattern involved an inverse-S function that crossed the 45° line somewhere between 0.3 and 0.4, overweighting probabilities lower than the crossing point and underweighting higher probabilities.

Such functions can readily produce the kinds of inequalities reported in Table A1 in online Appendix A which so resoundingly rejected H2. However, a function that crosses in the region of 0.3–0.4 will typically involve smaller deviations above the 45° line for lower probabilities than the deviations below the 45° line for the complementary higher probabilities it. This in turn would entail the sums of the PEs of the three paired-item claims tending to fall below 200% to a greater extent than the sums of the PEs of the three single-item claims exceed 100%. But that is not the typical pattern reported in Table 3: of the eight categories, five show a (considerably) larger deviation for the single-item claims, while in two cases the deviations are roughly the same, with only COLOURS in T2 exhibiting a substantially larger deviation for the paired-item claims.

As a further check, we estimated the parameters of Goldstein and Einhorn (1987) and Prelec (1998) probability weighting functions for each individual separately. In order to do this, we took each individual’s responses to Question 9 as indicative of their best estimates of the probabilities that each claim would pay £20 when those estimates are constrained to sum to 1. (Recall that this question elicits estimates of the numbers of other participants who will put each item top and requires these estimates to sum to 10.) We then regressed individuals’ ps from Questions 11–16 on those ‘baseline’ probability estimates (from Question 9, which henceforth we denote by P_B). Specifically, Goldstein and Einhorn’s (1987) probability weighting function, adapted to our notation, was estimated according to (1):

$$p = \frac{\beta(P_B)^\alpha}{\beta(P_B)^\alpha + (1 - P_B)^\alpha}, \quad (1)$$

where α determines the curvature and shape of the function and β determines its elevation. The two parameters together define the position when the function crosses the 45° line.

The Prelec function is estimated according to (2):

$$p = \exp[-\beta(-\ln P_B)^\alpha], \quad (2)$$

where, similarly to (1), α and β determine the curvature and elevation, respectively.

As with the functions fitted by Gonzalez and Wu (1999) and by Abdellaoui *et al.* (2011), there was considerable variability between individuals – see Tables A2 and A3 in online Appendix A for lists of the shapes of the fitted curves and the points (where applicable) at which they crossed the 45° line. The median (mean) crossing points of the inverse-S shaped curves are 0.548 (0.561) for the Goldstein and Einhorn specification and 0.533 (0.548) when the Prelec formulation is used. This represents a marked difference from the ‘typical’ functions that cross the 45° degree line somewhere between 0.3 and 0.4.

Abdellaoui *et al.* (2011) argued that the difference between the fitted curve under uncertainty and the typical curve under risk can be regarded as a measure of ambiguity attitude. In their data, the curves for the uncertainty treatments generally had lower elevation than their risk counterparts, which they took to indicate aversion to ambiguity. But our data clearly exhibit the opposite difference, producing curves with typically higher elevation. If our data were interpreted in terms of some attitude to ambiguity, they would indicate a predominance of ambiguity seeking. However, as we shall argue, there are other interpretations of the data that may be more persuasive.

4.2. Imprecision and Asymmetric Noise

A second way of explaining an inverse-S relationship involves no formal transformation function at all but, instead, considers the possibility that, when people are uncertain about the probabilities, their judgments may exhibit imprecision which manifests as asymmetric noise (MacCrimmon and Smith, 1986; Butler and Loomes, 2007). The key idea is that if an individual’s ‘true’ value/probability lies towards the lower end of some feasible range, imprecision about preference/judgment allows for more and larger deviations above that true value/probability than below it; and likewise if the true figure is towards the upper end of some feasible range, imprecision allows for more and larger deviations below it. So when a particular estimate from Question 9 is less than 5, implying a P_B below 0.5, there is more scope in the PE tasks for any imprecision to produce deviations above rather than below P_B , consistent with a tendency towards overweighting lower probabilities. Likewise, when the P_B inferred from Question 9 is above 0.5, there is more scope for deviations below rather than above, resulting on average in PEs lower than the P_B estimates would entail.

Such an account produces a relationship between the ps and their respective P_Bs which is consistent with an inverse-S pattern but which is more symmetrical, so that a fitted curve would be more likely to cut the 45° line in the vicinity of 0.5 than in the

0.3–0.4 region regarded as typical of risk-based probability transformation functions. A greater elevation of this kind would seem to provide a better fit with the data than can be achieved by the standard probability weighting function.

However, although noisy regression towards the mean may be part of the story, closer inspection of our data causes us to have some reservations about the adequacy of this explanation alone, in two respects. First, on the basis of such an account, one would generally expect that the more extreme the P_B , the more room there would be for a difference between it and the corresponding p ; but in fact there is no significant correlation of that kind. Second, it should not matter whether the probability inferred from Question 9 responses relates to a single-item claim or a paired-item claim; but in our data, this does appear to make a difference.

To examine whether there is some significant structural difference in the relationship between p and P_B in single-item and in paired-item cases, we ran a linear regression of p on P_B where we included both shift and slope dummies for the single-item data and obtained the results shown in columns (2)–(4) in Table 5. As a robustness check, we ran the regression again, but this time using the more fine-grained estimates of P_B taken from Question 10 responses: these results are shown in columns (5)–(7).

Results from both types of question show that there is a highly significant intercept and a positive slope significantly less than 1, consistent with $p > P_B$ for lower P_B and $p < P_B$ for higher P_B in the paired-item cases. There is no significant slope dummy, but the intercept increment of 0.04 in column (2) of Table 5 is significant at the 5% level and its counterpart of 0.054 in column (5) is significant at the 1% level, suggesting that, for any P_B , single-item based p s are 4% or 5% higher than their pair-based counterparts.

Such a structural difference is inconsistent with both of the candidate explanations discussed so far. However, there is a psychological model of probability judgment that might be better able to accommodate this feature of the data.⁷

Table 5

Results of the Regression with p as Dependent Variable and P_B as Explanatory Variable, with Intercept and Slope Dummies

| | P_B based on Question 9 | | | P_B based on Question 10 | | |
|-----------------|---------------------------|-------|-------------|----------------------------|-------|-------------|
| | Estimate | SE | t-statistic | Estimate | SE | t-statistic |
| Intercept | 0.131 | 0.014 | 9.335*** | 0.105 | 0.016 | 6.666*** |
| Slope | 0.710 | 0.020 | 14.811*** | 0.750 | 0.022 | 11.297*** |
| Intercept dummy | 0.040 | 0.016 | 2.450* | 0.054 | 0.018 | 2.965** |
| Slope dummy | 0.024 | 0.028 | 0.884 | 0.022 | 0.031 | 0.704 |

Notes. ***Significant at 0.001; **significant at 0.01; *significant at 0.05. Table 5 shows that Question 9 and Question 10 result in similar estimates of the relationship between p and P_B , with both exhibiting a significant structural difference between single items and pairs of items.

⁷ Although the formal model proposed by Abdellaoui *et al.* (2011, pp. 716–7) does not distinguish between the probability based on a single event and the same probability based on a union of more than one event, they note the possibility of what they refer to as ‘perceptual biases’ and point to the possible role of psychological effects such as those discussed in the next Section. Baillon (2008, pp. 83–4) makes a similar point, and also refers to ‘event-splitting effects’ (Starmer and Sugden, 1993; Humphrey, 1995) as a possible explanation of some indications in his results of departures from what we have called H2.

4.3. Support Theory

The patterns of data in our experiment bear some resemblance to subadditivity or ‘unpacking’ effects reported by psychologists in earlier probability judgment experiments and often explained by reference to Tversky and Koehler’s (1994) Support Theory (ST).

ST proposes that a ratio scale s can be assigned to any hypothesis about the occurrence of an event, where s represents the degree of support for that hypothesis. In the context of our experiment, $s(\text{Red})$ would be a measure of the degree of support for the hypothesis that the other player with whom one is matched would put Red top in the Question 1 ranking. Tversky and Koehler (1994, p. 549) suggested that ‘The support associated with a given hypothesis is interpreted as a measure of the strength of the evidence in favour of this hypothesis that is available to the judge. The support may be based upon objective data ... or on a subjective impression mediated by judgmental heuristics such as representativeness, availability, or anchoring and adjustment’. We suppose that in the context of our experiment, this translates into $s(\text{Red})$ being a measure of the number/strength of reasons that come to mind for Red being put top. Likewise, $s(\text{Blue})$ and $s(\text{Green})$ represent, respectively, the strengths of arguments for someone putting Blue or Green at the top of their ranking.

ST then entails that the judged probability that Red rather than Blue or Green is top-ranked when all three are explicitly identified at the same time can be expressed as $s(\text{Red})/[s(\text{Red}) + s(\text{Blue}) + s(\text{Green})]$. The corresponding judged probabilities for Blue and Green under the same circumstances are obtained by substituting each of $s(\text{Blue})$ and $s(\text{Green})$ for $s(\text{Red})$ in the numerator of that expression. In terms of ST, our P_B values obtained from the Question 9 and Question 10 tasks might be regarded as the best estimates of these three judged probabilities, since the three hypotheses concerning Red, Blue and Green were presented explicitly and simultaneously in those tasks. Put more generally in terms of the notation we have used to formulate our hypotheses:

$$P_B(X) = s(X)/[s(X) + s(Y) + s(Z)]. \quad (3)$$

ST goes on to propose that when two (or more) hypotheses are implicit rather than explicit, they have less support than the sum of their separate $s(\cdot)$ s: that is, if a hypothesis A is an implicit composite of two mutually exclusive explicit hypotheses A_1 and A_2 , then $s(A) \leq s(A_1) + s(A_2)$. In the context of our three-item design, ‘Not X’ is an implicit composite of the other two mutually exclusive states Y and Z. More formally:

$$s(\text{Not } X) \leq s(Y) + s(Z). \quad (4)$$

In terms of our tasks, the elicitation of the probability equivalent for a single item such as Red was worded as ‘You receive £20 if the other participant has put Red as their top ranking. You receive nothing otherwise’. The judgment of $p(\text{Red})$ can be derived by comparing the support for the hypothesis ‘the other person put Red top’ with the alternative ‘the other person’s top choice was Not Red’. Thus, $p(\text{Red}) = s(\text{Red})/[s(\text{Red}) + s(\text{Not Red})]$. More generally:

$$p(X) = s(X)/[s(X) + s(\text{Not } X)]. \quad (5)$$

In cases where $s(\text{Not } X) < s(Y) + s(Z)$, the denominator in (5) will be smaller than the denominator in (3). So for single-item cases, we should expect $p(i) > P_B(i)$ for $I = X, Y, Z$. And as Table 2 shows, that is what we find for the great majority of items: in 22 cases out of 24, the mean $p(i)$ s (given in brackets in the Questions 11–13 column) are higher than their $P_B(i)$ counterparts in both Question 9 (when converted to percentages) and in Question 10 (where the responses were elicited in percentage form). The only exceptions are Diamond and Car.

Subsequent development of ST by Rottenstreich and Tversky (1997, p. 407) proposed an extension of the model to give (when translated into our notation):

$$s(\text{Not } X) \leq [s(Y) \vee s(Z)] \leq s(Y) + s(Z), \quad (6)$$

where $[s(Y) \vee s(Z)]$ signifies ‘either Y or Z ’. The left-hand inequality represents implicit subadditivity while the right-hand inequality is referred to as explicit subadditivity. When this extension of ST is applied to our data, it entails that the p for any paired-item event (e.g. ‘either Blue or Green’ – which was the wording used in Questions 14–16) will be less than the sum of the P_B responses for those two colours as elicited via Questions 9 or 10. And this is what we find in 18 of the 24 pairs. The chances of such an asymmetry occurring by chance are approximately 1% ($p = 0.011$ in a one-tailed exact binomial test).

The relationship between implicit and explicit subadditivity, as expressed in the previous paragraph, also provides an account of the regression results reported in Table 5, where a significant intercept increment suggested that, noise aside, for a given P_B the corresponding p for a single-item claim would be 4–5% higher than its pair-based counterpart. To give an illustrative example, let us offset the inequalities in (6) by counterbalancing weights to give the following (7):

$$1.2s(\text{Not } X) = 1.1[s(Y) \vee s(Z)] = s(Y) + s(Z). \quad (7)$$

Suppose there are three items X, Y and Z whose $s(\cdot)$ – and therefore P_B – values are, respectively, 0.5, 0.3 and 0.2, so that the P_B for the single item X is equal to the sum of the P_B s for the two items Y and Z , with both being 0.5.

Now compute the probability equivalent for X . From (7), since $s(Y) + s(Z) = 0.5$, $s(\text{Not } X) = 0.5/1.2 = 0.4167$; thus $p(X) = s(X)/[s(X) + s(\text{Not } X)] = 0.5/0.9167 = 0.5455$; and hence $p(X) > P_B(X)$.

Next, compute $p(Y \cup Z)$. From (7), $s(Y \cup Z) = [s(Y) \vee s(Z)] = [s(Y) + s(Z)]/1.1 = 0.4545$, so that $p(Y \cup Z) = s(Y \cup Z)/[s(Y \cup Z) + s(X)] = 0.4545/0.9545 = 0.4762$. So $p(Y \cup Z) < [P_B(Y) + P_B(Z)]$.

The difference between $p(X)$ and $p(Y \cup Z)$ is close to the difference entailed by the regression reported in Table 5: were we to substitute $P_B = 0.5$ into that regression, it would give a p of 0.538 for a single item and a p of 0.486 for a pair of items. Of course, this is only an illustrative example but it shows how ST might account for the patterns we have observed.

5. Conclusion

We set out to elicit individuals’ subjective probabilities in a setting where there is uncertainty about others’ behaviour. Our experimental environment was therefore

rather different from the various ‘balls-in-an-urn’ designs which have often been used to investigate attitudes to ambiguity but it was arguably more similar to many real-world interactive environments. Our results were also rather different from those found in many traditional designs: although they were highly systematic, the behaviour of the great majority of our participants could not be explained in terms of any consistent ‘attitude to ambiguity’.

However, as indicated in the previous Section, there are alternative possible explanations for the data which require no invocation of any ‘attitude to ambiguity’ *per se*. What may be (much) more important is the process by which individuals arrive at their probability judgments, with this process being liable to systematic influences of the kind entailed by Support Theory. This does not exclude the possibility that those probability judgments are also imprecise and subject to asymmetric noise that may pull them somewhat towards 50%.

The net effect is quite clear: the basic assumptions of Subjective Expected Utility Theory, as represented in hypotheses H1–H3, fail systematically and significantly in the strategic setting we investigated. Although theorists have developed an impressive array of models designed to accommodate data from Ellsberg-urn experiments or other individual decision tasks involving some degree of ambiguity, such models do not appear to be sufficiently rich to capture the patterns we find when uncertainty arises from the strategic choices of other human beings.

University of Warwick

Additional Supporting Information may be found in the online version of this article:

Appendix A. Examples of Experimental Instructions.

Data S1.

References

- Abdellaoui, M., Baillon, A., Placido, L. and Wakker, P. (2011). ‘The rich domain of uncertainty: source functions and their experimental implementation’, *American Economic Review*, vol. 101(2), pp. 695–723.
- Abdellaoui, M., Klibanoff, P. and Placido, L. (2015). ‘Experiments on compound risk in relation to simple risk and to ambiguity’, *Management Science*, vol. 61(6), pp. 1306–22.
- Baillon, A. (2008). ‘Eliciting subjective probabilities through exchangeable events: an advantage and a limitation’, *Decision Analysis*, vol. 5(2), pp. 76–87.
- Bardsley, N., Mehta, J., Starmer, C. and Sugden, R. (2010). ‘Explaining focal points: cognitive hierarchy theory versus team reasoning’, *ECONOMIC JOURNAL*, vol. 120(543), pp. 40–79.
- Becker, G., DeGroot, M. and Marschak, J. (1964). ‘Measuring utility by a single-response sequential method’, *Behavioural Sciences*, vol. 3(3), pp. 226–32.
- Bernasconi, M. and Loomes, G. (1992). ‘Failures of the reduction principle in an Ellsberg-type problem’, *Theory and Decision*, vol. 32(4), pp. 77–100.
- Binmore, K., Stewart, L. and Voorhoeve, A. (2012). ‘How much ambiguity aversion?’, *Journal of Risk and Uncertainty*, vol. 45(3), pp. 215–38.
- Brenner, L., Koehler, D. and Rottenstreich, Y. (2002). ‘Remarks on support theory: recent advances and future directions’, in (T. Gilovich, D. Griffin and D. Kahneman, eds.), *Hewistics and Biases: The Psychology of Intuitive Judgement*, pp. 489–509. Cambridge: Cambridge University Press.
- Butler, D.J. and Loomes, G. (2007). ‘Imprecision as an account of the preference reversal phenomenon’, *American Economic Review*, vol. 97(1), pp. 277–97.
- Ellsberg, D. (1961). ‘Risk, ambiguity and the Savage axioms’, *Quarterly Journal of Economics*, vol. 75(4), pp. 643–69.

- Goldstein, W.M. and Einhorn, H.J. (1987). 'Expression theory and the preference reversal phenomena', *Psychological Review*, vol. 94(2), pp. 236–54.
- Gonzalez, R. and Wu, G. (1999). 'On the shape of the probability weighting function', *Cognitive Psychology*, vol. 38(1), pp. 129–66.
- Halevy, Y. (2007). 'Ellsberg revisited: an experimental study', *Econometrica*, vol. 75(2), pp. 503–36.
- Hey, J., Lotito, G. and Maffioletti, A. (2010). 'The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity', *Journal of Risk and Uncertainty*, vol. 41(2), pp. 81–111.
- Humphrey, S. (1995). 'Regret aversion or event splitting effects? More evidence under risk and uncertainty', *Journal of Risk and Uncertainty*, vol. 11(3), pp. 263–74.
- Kahneman, D. and Tversky, A. (1979). 'Prospect theory: an analysis of decision under risk', *Econometrica*, vol. 47(2), pp. 263–91.
- Loomes, G. and Pogrebna, G. (2014). 'Measuring individual risk attitudes when preferences are imprecise', *ECONOMIC JOURNAL*, vol. 124(576), pp. 569–93.
- Maafi, H. (2011). 'Preference reversals under ambiguity', *Management Science*, vol. 57(11), pp. 2054–66.
- MacCrimmon, K. and Smith, M. (1986). 'Imprecise equivalences: preference reversals in money and probability', Working Paper No. 1211, University of British Columbia.
- Machina, M. and Schmeidler, D. (1992). 'A more robust definition of subjective probability', *Econometrica*, vol. 60(4), pp. 745–80.
- Mehta, J., Starmer, C. and Sugden, R. (1994). 'Focal points in pure coordination games: an experimental investigation', *Theory and Decision*, vol. 36(2), pp. 163–85.
- Prelec, D. (1998). 'The probability weighting function', *Econometrica*, vol. 66(3), pp. 497–527.
- Quiggin, J. (1982). 'A theory of anticipated utility', *Journal of Economic Behavior and Organization*, vol. 3(4), pp. 323–43.
- Rottenstreich, Y. and Tversky, A. (1997). 'Unpacking, repacking and anchoring: advances in support theory', *Psychological Review*, vol. 104(2), pp. 406–15.
- Savage, L. (1954). *The Foundations of Statistics*, New York: Wiley. Revised and enlarged edition (1972), New York: Dover.
- Schelling, T.C. (1980). *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.
- Starmer, C. and Sugden, R. (1993). 'Testing for juxtaposition and event splitting effects', *Journal of Risk and Uncertainty*, vol. 6(3), pp. 235–54.
- Tversky, A. and Kahneman, D. (1992). 'Advances in prospect theory: cumulative representation of uncertainty', *Journal of Risk and Uncertainty*, vol. 5(4), pp. 297–323.
- Tversky, A. and Koehler, D.J. (1994). 'Support theory: a nonextensional representation of subjective probability', *Psychological Review*, vol. 101(4), p. 547.