# The explanatory and predictive power of non two-stage-probability theories of decision making under ambiguity

**John D. Hey · Noemi Pace**

**Abstract** Representing ambiguity in the laboratory using a Bingo Blower (which is transparent and not manipulable) and asking the subjects a series of allocation questions, we obtain data from which we can estimate by maximum likelihood methods (with explicit assumptions about the errors made by the subjects) a significant subset of particular parameterisations of the empirically relevant models of behaviour under ambiguity, and compare their relative explanatory and predictive abilities. Our results suggest that not all recent models of behaviour represent a major improvement in explanatory and predictive power, particularly the more theoretically sophisticated ones.

This paper compares the descriptive and predictive performance of particular parameterisations of five non two-stage-probability models of behaviour under

J. D. Hey (✉)
Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, UK
e-mail: john.hey@york.ac.uk; jdh1@york.ac.uk

N. Pace
Department of Economics, Universy Ca' Foscari of Venice, Cannaregio 873, 30121 Venezia, Italy
e-mail: n.pace@unive.it

ambiguity: Subjective Expected Utility, Choquet Expected Utility, Alpha Expected Utility, Vector Expected Utility and the Contraction Model.

The past decade has seen intense theoretical work in the modelling of behaviour under ambiguity. Now it is the time for the experimentalists to investigate the empirical validity of these theories. That is the primary purpose of this paper. Specifically, we complement a growing experimental literature, and, in particular, add to the work of Abdellaoui et al. (2011), Halevy (2007), Ahn et al. (2010) and Hey et al. (2010), though our detailed objectives, methods and results differ in various respects from theirs.

In essence, all these papers (and others) are aimed at the same fundamental objective: to discover which of the many theories of behaviour under ambiguity are empirically most appealing. However our work differs from these earlier works in terms of: (1) the representation of ambiguity (except for Hey et al. 2010); (2) the experimental design (except for Ahn et al. 2010); (3) the theories being explored; and (4) the econometric methods (except for part of Hey et al. 2010).

Ambiguity is represented in different ways in the experiments on which these different papers were based. Ambiguity is understood as a situation in which probabilities do not exist or the decision-maker does not know the actual probabilities. Both Halevy (2007) and Abdellaoui et al. (2011) use as one of their representations the traditional 'Ellsberg Urn': subjects are told what objects may be in the urn but are not told the quantities of each object, so that the probability of drawing any particular object cannot be known by the subject. Abdellaoui et al. (2011), given that their objective is to examine the impact of different *sources* of ambiguity, consider various sources (changes in the French Stock Index, the temperature in Paris, and the temperature at some randomly drawn remote country—all on a particular day). Ahn et al. (2010)'s representation consists of not telling the subjects what the precise probability of two of the three possible outcomes was; this is a sort of continuous 'Ellsberg Urn'. In contrast, Hey et al. (2010) used a Bingo Blower. This is also what we used in the experiment reported on in this paper. The Blower enables us to carry out two treatments which we feel have different amounts of ambiguity.

The papers by Hey et al. (2010) and Abdellaoui et al. (2011) use the 'traditional' form of experimental question: pairwise choices, while Halevy (2007) uses reservation price questions. In contrast, Ahn et al. (2010) use the *allocation* type of question pioneered originally by Loomes (1991), revived by Andreoni and Miller (2002) in a social choice context, and later by Choi et al. (2007) in a risky choice context. In this paper we use allocation problems, which are possibly[1] more informative than pairwise choice questions and reservation price questions. In this respect the comments by Wilcox (2007) as to the informative nature of experimental data in general, and pairwise choice questions in particular, should be noted. He discusses the necessity for considering the number and type of questions when designing an experiment.

---

[1]Though we should admit that the issue of the 'best' way to elicit preferences (whether by pairwise choices, Holt-Laury prices lists, the Becker-DeGroot-Marschak mechanism, by allocation questions or by some other method) is an open one.

The set of theories of behaviour under ambiguity is now very large. The earlier theories—such as maximin, which do not incorporate a preference functional—do not seem to have strong empirical support (see Hey et al. 2010); we do not consider them here. Of the remaining theories, one can make a broad distinction between the set of theories that uses second-order probabilities and the set that does not. For example, if there are $I$ possible events $i = 1, 2, ..., I$, but the probabilities of them are not known, those theories that use second-order probabilities assume that the decision-maker works on the basis that there is a set of $J$ possible values for these probabilities, with the $j^{th}$ set taking values $p_{1j}, p_{2j}, ..., p_{Ij}$ with the probability of the $j^{th}$ set being true given by $\pi_j$, $j = 1, 2, ..., J$. In contrast, the set of theories that does not use second-order probabilities may assume that the $p_i$ may take a range of values in the decision-maker's mind, but he or she does not attach probabilities to these possible values. We restrict attention here to this second set (non two-stage-probability models). This is for three reasons: the way we represent ambiguity in the laboratory; the complexity of the resulting models in the two-stage-probability set; and problems with distinguishability of the underlying preference functionals (because of the large number of parameters). In contrast, Halevy uses two-stage-probability models because his experimental design effectively makes such models appropriate. One could also argue that the same applies to the Ahn et al. experiment: there they have three possible outcomes 1, 2 and 3. Subjects are told $p_2$ but they are not told anything about $p_1$ and $p_3$ (except that they obey the usual probability rules). However, if subjects had read footnote 4 of their paper[2] then a two-stage-probability representation would have been natural.

Ahn et al. (2010) make an important distinction amongst the various specifications of behaviour under ambiguity: between those specifications which are *smooth* and those that are *kinked*.[3] Essentially this distinction consists of whether preference depends upon the *ordering* of the outcomes: in Expected Utility theory this is not the case and hence this is a smooth specification; in contrast, Choquet Expected Utility and, as one can tell from its name, Rank Dependent Expected Utility, are kinked. Ahn et al. (2010) do not estimate particular preference functionals but rather two general specifications—one smooth and one kinked. They note that the smooth specification "can be derived from" Recursive Expected Utility (REU, which is a two-stage-probability model—see Segal 1987), while the kinked specification "... can be derived as a special case of a variety of utility models: MEU, CEU, Contraction Expected Utility, and $\alpha$-MEU".[4] We note two things: first that the smooth specification does not come *only* from REU (indeed it comes from several other models, such as the Variational model of Maccheroni et al. 2006); and secondly, but perhaps more importantly, Ahn et al. (2010) do not estimate any preference functionals that come

---

[2]Which reads "In practice, the probability of one of the 'ambiguous' states was drawn from the uniform distribution over [0,2/3]. This distribution was not announced to the subjects."

[3]One way of thinking about this distinction is through indifference curves in outcome space. With the smooth models indifference curves are smoothly continuous, in contrast with kinked models, where there is a kink (a change in the slope) in the indifference curves along the certainty line.

[4](Our note) MEU, CEU and $\alpha$-MEU are respectively MaxMin Expected Utility, Choquet Expected Utility, and Alpha Expected Utility (all of which we consider specifically later).

*specifically* from the models that they mention. So they do not test *directly* any of the recent theories of behaviour under ambiguity.

Abdellaoui et al. (2011) effectively investigate only one model—essentially Rank Dependent Expected Utility (RDEU) theory. This has two key elements, a utility function, which they take to be CRRA (Constant Relative Risk Averse), and a (probability) weighting function, which they take, in the 'Ellsberg Urn' part of the experiment, to be of the Prelec form: $w(p) = (\exp(-(-\ln(p))^{\alpha}))^{\beta}$. It may seem a bit odd using probabilities in a study of ambiguity, but these probabilities are the true probabilities, which, of course, the experimenters know even if the subjects don't. Using these functional specifications, RDEU is a special case[5] of Choquet Expected Utility, which we estimate. In the 'Natural Uncertainties' part of the experiment they do not assume any particular form for the weighting function, so RDEU in this context is precisely Choquet—which we estimate.

There are significant econometric differences between these various papers. First, we (like some others) carried out extensive pre-experimental simulations to ensure that we had a sufficient number and an appropriate set of problems to ask the subjects; some experiments have rather few problems and thus lack power to discriminate between the theories. Second, the estimation methods vary. Underlying any particular chosen estimation method, there is an assumption about the stochastic specification of the model. Sometimes this is tacit; we feel it should be explicit, particularly as there is an obvious source for the stochastic component of the data—if one is estimating subject by subject (which is the case in all these papers) this comes either from randomness in preferences or from errors made by the subjects. While random preference models are popular with some theorists, they are difficult to parameterise and rarely applied to experimental data, so we follow the majority in assuming that the noise, the stochastic component, comes from errors made by the subjects. We explicitly include a story of such mistakes. In the results we report here, we assume a beta distribution for the error component, which seems the natural conjugate of our CRRA utility functional specification (though we have also investigated a normal distribution and a CARA utility function). We estimate the various preference functionals with the stochastic specification specifically built in to the estimation (like Andersen et al. 2009).

We also go one step further than some of these papers. Believing that economics is all about predicting, rather than just explaining, we compare our different models by seeing how good they are at predicting. For the importance of this, see Wilcox (2007, 2011).

In summary: we represent ambiguity in the laboratory in an open and non-manipulable manner; we ask a set of allocation problems to the subjects (obviously with an appropriate incentive mechanism) chosen after extensive simulations; we use maximum likelihood estimation, with a carefully-chosen stochastic specification, to estimate a significant sub-set of the empirically relevant theories of behaviour under

---

[5]If they had estimated the weighting function at all points, rather than estimating the parameters of the particular functional form, it would have been precisely Choquet.

ambiguity and to compare their relative goodness of fit; finally we compare the various theories in terms of their predictive ability.

The rest of the paper is organised as follows. In the next section we give a brief overview of the theories that we are going to fit to our data. The following section describes the way that our data was generated in our experiment. We then relate what we do to the literature, a part of which we have discussed in this introduction, and give more detail about what others have done. A section describing the technicalities underlying our analysis then follows, after which we present our results. We then conclude.

## 1 Theories under investigation

This section discusses the theories of decision-making under ambiguity that we investigate. We confine our attention to those theories in which there is an explicit preference functional, and hence we exclude earlier theories which proceed directly to a decision rule.[6] In all the theories under consideration here, the decision maker is viewed, in any decision problem, as if maximising the value of some preference functional. The theories differ in the 'stories' that they are telling. The familiar (1) Subjective Expected Utility (SEU) model postulates the decision-maker as being able to attach *subjective probabilities* to the various possible events. In contrast, (2) the Choquet Expected Utility (CEU) model, usually nowadays accredited to Schmeidler (1989), postulates that the agent's beliefs cannot be characterised by additive probabilities but by non-additive *capacities*; (3) the Alpha Expected Utility (AEU) model of Ghirardato et al. (2004) models the agent as not being able to specify *unique* probabilities but instead a set of possible probabilities (though not attaching probabilities to the members of this set); (4) the Vector Expected Utility (VEU) model of Siniscalchi (2009) sees the decision-maker assessing an uncertain prospect with a baseline expected utility evaluation and an "adjustment that reflects the individual's perception of ambiguity and her attitude toward it" (Siniscalchi 2009, page 1). Finally (5) the Contraction Model (COM) of Gajdos et al. (2008) is similar to the AEU model, in terms of specifying a set of possible probabilities, but processes the expected utilities over this set in a different way. We note that SEU is a 'smooth' specification in the sense used above, while all the rest are kinked specifications. We give an overview of these theories below. We restrict attention in both the overview and the detail to decision problems with at most three events—which was the case in our experiment. Call these events $E_1$, $E_2$ and $E_3$. For the decision maker, each event will be associated with an outcome which consists of an amount of money. For some of the theories—those with a 'rank-dependent' flavouring—the ordering of the outcomes will be crucial. We denote the various possible utilities of the decision-maker by $u_{(1)}$, $u_{(2)}$ and $u_{(3)}$ where we assume that $u_{(1)} \geq u_{(2)} \geq u_{(3)}$. Denote by $E_{(i)}$ the event that leads to outcome $u_{(i)}$ *(i=1,2,3)*. So the set $\{E_{(1)}, E_{(2)}, E_{(3)}\}$ has the same

---

[6] Such as, for example, MaxMin (in which the decision-maker looks at the worst that can happen and makes that as good as possible) and MaxMax (in which the decision-maker looks at the best that can happen and makes that as good as possible). See Hey et al. (2010) for the empirical evidence against such theories.

elements as the set $\{E_1, E_2, E_3\}$ though not necessarily in the same order. Note crucially the set $\{E_{(1)}, E_{(2)}, E_{(3)}\}$ may vary from problem to problem depending upon the subject's decisions.

Our method is that of estimation and prediction. We need, therefore, to estimate parameters. All of the theories we investigate include a utility function; this has to be specified and parameterised. Of the five theories under investigation, two do not contain any functions other than the utility function. These are Subjective Expected Utility theory (which uses subjective probabilities), and Choquet Expected Utility (which uses capacities). The other three theories, as we shall show, contain functions and sets each of which needs to be specified—otherwise estimation cannot proceed. As these three theories are formulated, there is an infinite number of possible parameterisations, and the theories do not provide any guidance, except for the properties of the functions and sets. We adopt what seems to us simple specifications in all cases, though it is clearly possible that different specifications would give different results. We do not claim that we provide a general test of the various models.

## 1.1 Subjective expected utility theory

The preference functional for SEU is given by

$$SEU = \sum_{i=1}^{3} p_{(i)} u_{(i)} \tag{1}$$

where $p_{(i)}$ is the subjective probability that event $E_{(i)}$ occurs, so that $p_{(i)} = \text{Prob}(E_{(i)})$ for all $i$, and, of course $p_{(1)} + p_{(2)} + p_{(3)} = 1$.

## 1.2 Choquet expected utility theory

According to Schmeidler (1989), the Choquet Expected Utility of a lottery is given by

$$CEU = \sum_{i=1}^{3} \bar{w}_i u_{(i)} \tag{2}$$

where the $\bar{w}$'s are weights that depend on nonadditive *capacities* $w$ that satisfy the normalisation conditions and monotonicity (with respect to set inclusion). In the context of our experiment, a CEU subject works with six nonadditive capacities $w_{E_1}, w_{E_2}, w_{E_3}, w_{E_2 \cup E_3}, w_{E_3 \cup E_1}$ and $w_{E_1 \cup E_2}$ referring to the three events and their pairwise unions. From these we can derive the set of capacities relevant to any decision problem $\{w_{E_{(1)}}, w_{E_{(2)}}, w_{E_{(3)}}, w_{E_{(2)} \cup E_{(3)}}, w_{E_{(3)} \cup E_{(1)}}, w_{E_{(1)} \cup E_{(2)}}\}$ which are a permutation of the set $\{w_{E_1}, w_{E_2}, w_{E_3}, w_{E_2 \cup E_3}, w_{E_3 \cup E_1}, w_{E_1 \cup E_2}\}$. Crucially, the weights $\bar{w}$ depend upon the ordering of the outcomes:

$$\begin{aligned} \bar{w}_1 &= w_{E_{(1)}} \\ \bar{w}_2 &= w_{E_{(1)} \cup E_{(2)}} - w_{E_{(1)}} \\ \bar{w}_3 &= 1 - w_{E_{(1)} \cup E_{(2)}} \end{aligned} \tag{3}$$

We note that the main difference between CEU and SEU consists in the additive probability measure being replaced by a nonadditive capacity measure. If the capacities are actually probabilities (that is, if $w_{E_2 \cup E_3} = w_{E_2} + w_{E_3}$, $w_{E_3 \cup E_1} = w_{E_1} + w_{E_3}$, $w_{E_1 \cup E_2} = w_{E_1} + w_{E_2}$ and $w_{E_1} + w_{E_2} + w_{E_3} = 1$ ) then Eq. 2 is equivalent to Eq. 1. We note that CEU is the same as Rank Dependent Expected Utility (see Wakker 2010 for a discussion of how this can be used as a model of decision-making under ambiguity) under an appropriate interpretation of that latter theory.[7] Similarly Cumulative Prospect Theory, with a fixed reference point, can be regarded in the same way as CEU.

### 1.3 Alpha expected utility theory

Alpha Expected Utility theory (AEU) was proposed by Ghirardato et al. (2004) as a generalization of the theory proposed in Gilboa and Schmeidler (1989). Ghirardato et al. (2004)'s model implies that, although the decision maker does not know the true probabilities, he or she acts as if he or she believes that the true probabilities lie within a continuous set $D$ of probabilities of different events. We can refer to each prior $p \in D$ as a "possible scenario" that the decision maker envisions. According to Ghirardato *et al*, the set $D$ of probabilities represents formally the ambiguity that the decision maker feels in the decision problem (they introduce the concept of "revealed ambiguity"). In other words, the size of the set $D$ measures the perception of ambiguity. The larger $D$ is, the more ambiguity the decision maker appears to perceive in the decision problem. In particular, no decision maker perceives less ambiguity than one who reveals a singleton set $D = \{p_1, p_2, p_3\}$. In this case the decision maker is a SEU maximiser with subjective probabilities $p_1$, $p_2$ and $p_3$.

According to Alpha Expected Utility Theory, decisions are made on the basis of a weighted average of the minimum expected utility over the set $D$ of probabilities and the maximum expected utility over this set:

$$AEU = \alpha \min_{p \in D} \sum_{i=1}^{3} p_i u_i + (1 - \alpha) \max_{p \in D} \sum_{i=1}^{3} p_i u_i \qquad (4)$$

The parameter $\alpha$ can be interpreted as an index of the ambiguity aversion of the decision maker. The larger $\alpha$ is, the larger is the weight the decision maker gives to the pessimistic evaluation given by $\min_{p \in D} \sum_{i=1}^{3} p_i u_i$.

In order to proceed to estimation we need to characterise the set $D$. The theory merely states some general characteristics (convexity and continuity) but no specific

---

[7]In the context of our experiment, where there are three outcomes and hence six capacities, then the relationship between the two theories is given by the following, where $p_1$, $p_2$, $p_3$ are the objective probabilities and $w(.)$ is the weighting function, and the capacities for CEU are as denoted above:

$w_{E_i} = w(p_i)$ for $i = 1, 2, 3$ and

$w_{E_j \cup E_k} = w(p_j + p_k)$ for $j \neq k \in 1, 2, 3$

set. We have to parameterise this set and we chose one of the simplest: that the set is defined by three lower bounds $\underline{p}_1$, $\underline{p}_2$ and $\underline{p}_3$ (where $\underline{p}_1 + \underline{p}_2 + \underline{p}_3 \leq 1$) plus the condition that every element in the set has $p_1 \geq \underline{p}_1$, $p_2 \geq \underline{p}_2$ and $p_3 \geq \underline{p}_3$. In addition, of course $p_1 + p_2 + p_3 = 1$ for each element in the set. These conditions imply that the set $D$ is a triangle properly within the Marschak-Machina Triangle. It reduces to a single point, and hence AEU reduces to SEU, if $\underline{p}_1 + \underline{p}_2 + \underline{p}_3 = 1$. There are obviously other specifications of the set $D$ that we could adopt—indeed there is an infinite number; for example, the set could be a circle within the Marschak-Machina triangle. Our results may well be sensitive to our specification, but without making such a specification, the theory is unestimatable.

## 1.4 Vector expected utility

Vector Expected Utility (VEU) theory has been recently proposed by Siniscalchi (2009). In this model, an uncertain prospect is assessed according to a baseline expected utility evaluation and an adjustment that reflects the individual's perception of ambiguity and his or her attitude toward it. This adjustment is itself a function of the exposure to distinct sources of ambiguity, and its variability.

The key elements of the VEU model are a baseline probability and a collection of random variables, or adjustment factors, which represent acts exposed to distinct ambiguity sources and also reflect complementarity between ambiguous events.

The VEU model can be formally defined as follows:

$$VEU = \sum_{i=1}^{3} p_i u_i + A\left( \left( \sum_{i=1}^{3} p_i \zeta_{ji} u_i \right)_{j=1,2,3} \right) \tag{5}$$

Here $p = (p_1, p_2, p_3)$ is the baseline prior; for $1 \leq j < 3$, each $\zeta_j = (\zeta_{j1}, .., \zeta_{j3})$ is an adjustment factor that satisfies $E_p[\zeta_j] = \sum_{i=1}^{3} p_i \zeta_{ji} = 0$; and A: $\mathbb{R}^n \to \mathbb{R}$ satisfies $A(0) = 0$ and $A(\phi) = A(-\phi)$. The function $A$ is an adjustment function that reflects attitudes towards ambiguity. We need to specify the function $A(.)$ and also the values of the $\zeta$.

We made the following assumptions.[8] For the baseline prior we have two parameters, $p_1$ and $p_2$ (because probabilities must sum up to one, thus $p_3 = 1 - p_1 - p_2$).

Now let us consider the adjustment factors. Since they have to integrate up to zero according to the theory, we take:

$$\zeta_{01} = \zeta_0(\omega_1) = \frac{1}{p_1} \zeta_{02} = \zeta_0(\omega_2) = -\frac{1}{p_2} \zeta_{03} = \zeta_0(\omega_3) = 0 \tag{6}$$

---

[8]These assumptions were made after private communication with Marciano Siniscalchi, though we do not imply that our modelling of the VEU model has his approval.

and similarly

$$\zeta_{11} = \zeta_1(\omega_1) = 0 \quad \zeta_{12} = \zeta_1(\omega_2) = \frac{1}{p_2} \quad \zeta_{13} = \zeta_{13}(\omega_3) = -\frac{1}{p_3} \tag{7}$$

This parsimonious specification does not restrict preferences in any *ex ante* way. The first factor captures ambiguity about the relative number of colour 1 balls and colour 2 balls, and the second reflects ambiguity about colour 2 versus colour 3 balls. The vectors $\zeta_0$ and $\zeta_1$ are linearly independent and satisfy $p \cdot \zeta_0 = p \cdot \zeta_1 = 0$.

Adopting this specification we find that the VEU functional is given by:

$$VEU = \sum_{i=1}^{3} p_i u_i + A(u_1 - u_2, u_2 - u_3) \tag{8}$$

We now need to choose a suitable function $A : \mathbb{R}^2 \rightarrow \mathbb{R}$. $A(0,0) = 0$ and $A(\phi_0, \phi_1) = A(-\phi_0, -\phi_1)$.

A relatively flexible specification is the following:

$$A(\phi_0, \phi_1) = \alpha[1 - (1 + |\phi_0|)^\rho] + \beta[1 - (1 + |\phi_1|)^\rho] \tag{9}$$

Plugging the specification of $A$ Eq. (9) in Eq. 8 we get the following objective function

$$VEU = \sum_{i=1}^{3} p_i u_i + \alpha[1 - (1 + |\phi_0|)^\rho] + \beta[1 - (1 + |\phi_1|)^\rho] \tag{10}$$

Since in our experiment, $|\phi_0| = |u_1 - u_2|$ and $|\phi_1| = |u_2 - u_3|$, Eq. 10 is equivalent to

$$VEU = \sum_{i=1}^{3} p_i u_i + \alpha[1 - (1 + |u_1 - u_2|)^\rho] + \beta[1 - (1 + |u_2 - u_3|)^\rho]$$

The parameter $\alpha$ and $\beta$ are likely to be the same since there is no reason why the ambiguity about the relative number of colour 1 balls and colour 2 balls should be different from the ambiguity about colour 2 versus colour 3. Therefore let us set $\alpha = \beta = \delta$. Assuming also that $\rho = 1$, the VEU objective function becomes

$$VEU = \sum_{i=1}^{3} p_i u_i - \delta(|u_1 - u_2| + |u_2 - u_3|) \tag{11}$$

This has intuitive appeal: decisions are made on the basis of expected utility 'corrected' for differences between the utilities of the various outcomes, weighted by a parameter $\delta$ that reflects the decision-maker's attitude toward ambiguity and the amount of ambiguity. We should note that the most general version of VEU (as specified in Eq. 5) has an infinite number of implementable forms, depending upon the

specification of the function $A(.)$ and $\zeta$, since the restrictions on $A(.)$ and $\zeta$ do not enable us to specify it precisely. It follows that Eq. 11 is just one of an infinite number of possible specifications of the VEU preference functional. When we refer to "estimating the VEU model" this restriction should be taken into account. But, as we noted above with AEU, we need to adopt a particular specification in order to proceed with estimation.

Finally, if the utilities are ordered (so that $u_1 \geq u_2 \geq u_3$) then this reduces to

$$VEU = (p_1 - \delta)u_1 + p_2u_2 + (p_3 + \delta)u_3 \qquad (12)$$

In order for the preference function to be increasing in the outcomes this requires that $\delta$ is sufficiently small, so that $p_1 - \delta > 0$. Our estimates always satisfied this condition.

## 1.5 The contraction model

Gajdos et al. (2008) proposed a model (the "Contraction Model" or COM) in which it is possible to compare acts under different objective information structures. According to this theory, preferences are given by

$$COM = \lambda \min_{p \in D} \sum_{i=1}^{3} p_iu_i + (1 - \lambda) \sum_{i=1}^{3} P_iu_i \qquad (13)$$

where $\lambda$ measures imprecision aversion and $P_1$, $P_2$, $P_3$ is a particular probability distribution in the set $D$ of possible distributions. It is what is called the 'Steiner Point' of the set—which is, in a particular sense, the 'centre' of the set. If we take the set $D$ of possible distributions as all points $(p_1, p_2, p_3)$ such that $p_1 + p_2 + p_3 = 1$ and $p_1 \geq \underline{p}_1, p_2 \geq \underline{p}_2, p_3 \geq \underline{p}_3$ then the Steiner point is the point $(P_1, P_2, P_3)$ where $P_i = \underline{p}_i + \left(1 - \underline{p}_1 - \underline{p}_2 - \underline{p}_3\right)/3$ for $i = 1, 2, 3$. We note that we have characterised this set $D$ (of possible probabilities) in the same way as we have done for the Alpha Expected Utility model—as a triangle properly within the Marschak-Machina Triangle—but there is no reason that the estimates of these lower bounds on the probabilities should be the same as the estimates of the lower bounds for the Alpha Expected Utility model. Nor is there any reason why the estimate of the parameter $\lambda$ in the Contraction Model should be the same as the estimate of $\alpha$ in the Alpha Expected Utility model.

In summary, each of our models (SEU, CEU, AEU, VEU and COM) implies a particular preference functional under our functional specifications, respectively Eqs. 1, 2, 4, 11 and 13. It is clear that these are different, except insofar as Eqs. 2, 4, 11 and 13 reduce to (1) (and hence CEU, AEU, VEU and COM reduce to SEU) when respectively, the CEU capacities are additive, the set D of probabilities in AEU consists of a single element, the parameter $\delta$ is zero, and the set $D$ of probabilities in COM consists of a single element. Given that Eqs. 2, 4, 11 and 13 are different it follows that the models are observationally distinguishable: different models imply

different preference functionals and hence different decisions. However this is not to deny that the function $A(.)$ and the $\zeta$ in VEU and the sets $D$ in AEU and COM could be such that they lead to the same preference functionals. But the crucial point is that our specifications of the different models imply different preference functionals and hence in principle are observationally distinguishable.

## 2 Our experimental design

As we have already noted, in our experiment ambiguity was implemented with a Bingo Blower and subjects were presented with a set of allocation problems, which were determined after a number of Monte Carlo simulations. We implemented two separate treatments, which we describe below. Different sets of subjects did the different treatments.

Subjects completed the experiment individually at screened computer terminals. They were given written instructions and then shown a PowerPoint presentation of the instructions. There was a Bingo Blower in action at the front of the laboratory throughout the experiment. The Bingo Blower is a rectangular-shaped, glass-sided, object some 3 feet high and 2 feet by 2 feet in horizontal section. Inside the glass walls are a set of balls in continuous motion being moved about by a jet of wind from a fan in the base. In addition, images of the Blower in action were projected via a video camera onto two big screens in the laboratory. Subjects were free at any stage to go close to the Blower to examine it as much as they wanted. All the balls inside the Blower can at all times be seen by people outside, but, unless the number of balls in the Blower is low, the number of balls of differing colours cannot be counted because they are continually moving around. Hence the information available is not sufficient to calculate objective probabilities. This ensures that, while objective probabilities do exist, the decision-makers cannot know them. In this way, we have created a situation of genuine ambiguity which eliminates the problem of possible suspicion,[9] the problem of directly using a second-order probability distribution, and does not necessitate the use of real events, therefore keeping the problem more similar to the original Ellsberg problem (Ellsberg 1961), though clearly we do not create a situation of 'total' or 'complete' ambiguity as Ellsberg tried to do. We note that a further advantage of this way of creating ambiguity in the laboratory is the fact that the information available is the same for all subjects. Hence there is no role for the so called 'comparative ignorance' (Fox and Tvesky 1995), and hence we can exclude such a factor as a possible explanation of behaviour.[10]

---

[9]We do not deny that some subjects could suspect that different coloured balls had different weights, but that could have been checked after the experiment. No subject asked for such a check.

[10]One criticism concerning the implementation of ambiguity in the lab using the Bingo Blower comes from Morone and Ozdemir (2012). The criticism consists of the observation that the ability of getting the right probabilities is subject specific; that is, subjects have different counting skills, or might have problems in the perception of colours. This criticism may be true but it is not clear how this could affect the validity of the Bingo Blower in generating ambiguity in the lab. Moreover since we analyse the data subject-by-subject, it is unimportant if different subjects have different perceptions of the amount of ambiguity.

This Bingo Blower played an important role in representing ambiguity and in providing incentives. Inside the Bingo Blower were balls of three different colours: pink, yellow and blue. The number of each colour depended on the treatment:

|        | Treatment 1 | Treatment 2 |
|--------|-------------|-------------|
| pink   | 2           | 8           |
| yellow | 3           | 12          |
| blue   | 5           | 20          |

In Treatment 1 the pink and yellow balls could almost certainly be counted, though one might not be sure of the number of blue balls; this was *the least ambiguous treatment*. In Treatment 2 the balls of each colour could not be counted; this was *the most ambiguous treatment*. Note that in this latter treatment subjects could get some idea of the relative numbers of balls of the different colours but not count the numbers precisely. It was reasonably clear that there were more blue balls than yellow, and more yellow than pink, though precise counting was difficult.

Sixty-six subjects completed Treatment 1 and sixty-three completed Treatment 2. In both treatments, subjects were presented with a total of 76 problems. Each of these asked them to allocate a given quantity of tokens between the colours. There were two *types* of problem. Type 1 asked them to allocate the tokens between *two* of the colours (with an explicit allocation of zero tokens to the third colour); Type 2 asked them to allocate the tokens between one of the three colours and the other two. In each problem subjects were told the *exchange rate* between tokens and money for each of the colours in the problem. Thus an allocation of tokens implied an allocation of money to two or three of the colours.

We provided an incentive for carefully choosing the allocations with the following payment scheme. We told subjects that, after answering all 76 problems, one of the problems would be chosen at random, and the subject's allocation to the two or three colours for that problem retrieved from the computer. At that point the subject and the experimenter went over to the Bingo Blower, and the subject tilted the tube to expel one ball. The colour of the ball, the problem picked at random and their answer to that problem determined their payment. To be precise: if the problem chosen was one of Type 1, then they would be paid the money implied by their allocation to the colour of the ball expelled; if it was the colour not mentioned in that problem they would be paid nothing; if the problem chosen was one of Type 2, then they would be paid the money implied by their allocation to the colour of the ball expelled. In addition they received a show-up fee of £5. The average total payment was £24.60. They filled in a brief questionnaire, were paid, signed a receipt and were free to go. A total of 129 subjects participated in the experiments, 40 of them at CESARE at LUISS in Rome (Italy) and the remaining 89 at EXEC at the University of York (UK), with the software and the instructions in Italian and English respectively. There were no differences in the behaviour of the two subject pools. In both York and Rome, subjects were recruited using the ORSEE (Greiner 2004) software

and the experiment was run using a purpose-written software written in Visual Basic 6.

## 3 Related experimental literature

Having described our experimental implementation and motivation we are now in a position to survey the relevant experimental literature in more detail. We confine ourselves mainly to recent contributions to the literature; earlier literature is surveyed in Camerer and Weber (1992) and Camerer (1995) while more recent mainly theoretical literature is surveyed by Etner et al. (2012).

Hey et al. (2010), using the same implementation of ambiguity in the laboratory as we use here, also with three possible outcomes, but asking a large number (162) of pairwise choice questions, examined the descriptive and predictive ability of twelve theories of behaviour under ambiguity: some old and not using a preference functional (proceeding directly to a decision rule) such as the original MaxMin and MaxMax; and some recent, such as the Alpha Expected Utility model. The findings were that the old simple models (those without a preference function) did not command empirical support, and that more modern models (such as Choquet) performed rather marginally better than simpler theories such as Subjective Expected Utility theory. Estimation of the preference functions was done using maximum likelihood techniques with the stochastic specification determined by a model of how subjects made errors in their pairwise choices.

Ahn et al. (2010) used allocation questions, like we do here, but implemented ambiguity by not telling the subjects the true objective probabilities of two of the three possible outcomes of the experiment. They did not look at the predictive ability of any models; neither did they examine the descriptive performance of any specific theory. Instead they examined two broad classes of functionals, smooth and kinked, which are special cases of various theoretical models that we specifically estimate. Econometrically they estimated, subject by subject, the risk-aversion parameter of an assumed Constant Absolute Risk Aversion utility function, and a second parameter measuring ambiguity aversion, using Non Linear Least Squares (NLLS), that is by minimising the sum of squared differences between actual allocations and the theoretically optimal allocations for those risk and ambiguity aversion coefficients.

Halevy (2007) implemented ambiguity in the laboratory using traditional Ellsberg Urns and asked reservation price questions. Because of the way that his 'Ellsberg Urns' were implemented, his set of models includes some models that we do not consider here, particularly two-stage-probability models such as Recursive Nonexpected Utility and Recursive Expected Utility. But we include some that he does not, making the two papers complementary. He used *reservation price* questions; we should describe and discuss these as they are an alternative to pairwise choice questions and to allocation questions. Essentially he wanted to know how much subjects value bets on various events. Let us consider a particular Ellsberg Urn and a particular colour. The subject is asked to imagine that he or she owns a bet which pays a certain amount of money ($2) if that coloured ball is drawn from that particular urn. Halevy wanted

to elicit the subject's reservation price for this bet; this reservation price tells us about the subject's preferences. Halevy used the Becker-DeGroot-Marschak mechanism: "the subject was asked to state a minimal price at which she was willing to sell the bet... The subject set the selling price by moving a lever on a scale between $0 and $2. Then a random number between $0 and $2 was generated by the computer. The random number was the "buying price" for the bet. If the buying price was higher than the reservation price that the subject stated, she was paid the buying price (and her payoff did not depend on the outcome of her bet). However, if the buying price was lower than the minimal selling price, the actual payment depended on the outcome of her bet". This BDM technique is well-known in the literature. However there are well-known problems[11] (see Karni and Safra 1987) with using this technique when preferences are *not* expected utility preferences—which, of course, is precisely the concern of that paper. Halevy did not use his data to estimate preference functionals and hence did not compare their descriptive and predictive power; instead he carried out an extensive set of tests of the various theories. This econometric procedure does not help to draw unique conclusions about the 'best' preference functional, even for individual subjects. Indeed Halevy concludes that his "...findings indicate that currently there is no unique theoretical model that universally captures ambiguity preferences".

Abdellaoui et al. (2011) investigated only Rank Dependent Expected Utility theory. They did not explicitly examine its descriptive (nor predictive) ability, being more concerned with the effect on the estimated utility and weighting functions of different *sources* of ambiguity. As we have already noted, they implemented ambiguity in the laboratory in two ways: in one part of the experiment using 8-colour 'Ellsberg Urns'; and in the other part using 'natural' events. They elicited certainty equivalents (or reservation prices) in order to infer preferences, not using the BDM mechanism, but instead using Holt-Laury price lists.[12] This mechanism might be a better way of eliciting certainty equivalents, even though the outcome does appear to be sensitive to the elements in the list—the number of them and their range.[13] The resulting certainty equivalents are a valuation, just like Halevy's reservation prices, even though they come from a set of pairwise choice questions. However econometrically it must be the case that the valuation resulting from a list with $n$ elements is less informative than $n$ independent pairwise choice questions. They estimated utility functions (assumed to be power or CRRA) "using nonlinear least squares estimation with the certainty equivalent as dependent variable"; similarly they estimated the weighting function by "minimising the quadratic distance".

Andersen et al. (2009) use a technique similar to that used by Ahn et al. (2010) in estimating two parameters (one a measure of risk aversion and the other a measure of

---

[11]There are also problems, though of a different nature, involved with using our Random Lottery Incentive mechanism. But see http://people.few.eur.nl/wakker/miscella/debates/randomlinc.htm

[12]In the Holt-Laury price lists subjects are presented with a set of pairwise choices arranged in a list. In each pair subjects are asked to choose between some ambiguous lottery and some certain amount of money. As one goes down the list, the certain amount increases. The subject's certainty equivalent is revealed by the point at which the subject switches from choosing the lottery to choosing the certain amount. See Holt and Laury (2002).

[13]See Andersen et al. (2006).

ambiguity) in a minimalist non-EU model. They comment that this minimalist model comes either from the Source-Dependent Risk Attitude model or the Uncertain Priors model; in our terminology it is a two-stage-probability model[14] that looks exactly like Recursive Expected Utility. The bottom line is the following: suppose that there are $I$ possible outcomes $i = 1, 2, ..., I$ with unknown probabilities. The decision-maker has a set of $J$ possible values for these probabilities; we denote the $j$'th possible value $p_{1j}, p_{2j}, ..., p_{Ij}$ and the decision-maker considers that the probability that this is the correct set is $\pi_j$. The preference function is the maximisation of

$$\sum_{j=1}^{J} \pi_j v \left[ \sum_{i=1}^{I} p_{ij} u(x_i) \right]$$

Note that there are two functions here: $u(.)$ which can be considered as a normal utility function, capturing attitude to risk, and $v(.)$ which can be considered as an ambiguity function. Note that if $v(y) = y$ then this model reduces to Expected Utility theory. It is the non-linearity of $v(.)$ which captures aversion to ambiguity. Andersen et al. (2009) assumed that both these functions are power functions—so that $u(x) = x^\alpha$ (a CRRA utility function) and $v(y) = y^\beta$. They estimated the two parameters $\alpha$ and $\beta$ using maximum likelihood techniques (with careful attention paid to the stochastic specification) and assumptions[15] about the $\pi$'s and $p$'s.

Before we conclude this section we should make some comments about a key methodological point: some papers *test* theories; some, like ours, *estimate* preference functionals. Both approaches investigate the empirical plausibility of theories. They are complements to each other. With the testing approach, one can isolate different components (axioms) of the theories, and determine which are valid and which are not. With the estimation approach, one can see how well models as a whole fit the data, and how well they predict. The latter is useful if one wants to explain and predict; the former is useful, particularly to theorists, in helping to decide how theories should be modified and adapted in light of empirical evidence.

## 4 Technical assumptions

Before proceeding to our estimates we need to make some technical assumptions. In particular we need to decide on our stochastic specification and the form of the utility function. These are interrelated decisions. Both are important, as Wilcox (2007, 2008 and 2011) makes clear, though our context is different from his as the experimental task in our experiment requires subjects to make a series of *allocations*, rather than to make a series of *pairwise choices*. This has implications for the stochastic structure,

---

[14]Chambers et al. (2010) also investigate a generic Multiple Priors model.

[15]It should be noted that the authors admit that the assumptions were quite strong and that they discuss the serious identification problems with two-stage-probability models.

but this, in turn, depends upon the assumed utility function: the two issues are inter-related.

We tried three different combinations of the form of the utility function and the stochastic specification, but we report just one here as the other two seemed to be empirically inferior. The combination that we report starts with a Constant Relative Risk Aversion (CRRA) utility function (like Abdellaoui et al. 2011 and Andersen et al. 2009), written in the form:

$$u(x) = \frac{x^{1-r} - 1}{1 - r} \text{ if } r \neq 1 \tag{14}$$
$$\ln(x) \text{ if } r = 1$$

For all the models we can write the objective of the subject in both types of problems as the maximisation of some function

$$w_1 u(e_1 x_1) + w_2 u(e_2 x_2)$$

subject to $x_1 + x_2 = m$, where the $w$'s and the $e$'s are defined appropriately. The solution, if $u(.)$ takes the form above, is:

$$x_1^* = \frac{e_1^{\frac{1}{r}-1} w_1^{\frac{1}{r}}}{e_1^{\frac{1}{r}-1} w_1^{\frac{1}{r}} + e_2^{\frac{1}{r}-1} w_2^{\frac{1}{r}}} m \tag{15}$$

$$x_2^* = \frac{e_2^{\frac{1}{r}-1} w_2^{\frac{1}{r}}}{e_1^{\frac{1}{r}-1} w_1^{\frac{1}{r}} + e_2^{\frac{1}{r}-1} w_2^{\frac{1}{r}}} m$$

In this case, the optimal allocations are bounded between 0 and $m$. Hence the proportions $x_1^*/m$ and $x_2^*/m$, are bounded between 0 and 1. This suggests a conjugate stochastic specification, which fits in naturally with the boundedness of these optimal proportions, namely a Beta distribution. Specifically we take the actual proportional allocation $x_1/m$ to have a beta distribution with parameters $x_1^*(s-1)/m$ and $x_2^*(s-1)/m$. This guarantees that the mean of $x_1$ is $x_1^*$ and its variance is $x_1^* x_2^*/s$. So the variance of $x_1$ is not constant but is zero at 0 and $m$ and reaches a maximum when $x_1^* = m/2$. It also follows (since $x_1^* + x_2^* = 1$) that $x_2$ has a beta distribution with parameters $x_1^*(s-1)$ and $x_2^*(s-1)$. Of course we may still observe actual proportional allocations equal to 0 and 1 because of the rounding of subjects' choices.

For the record, we note that we tried two other combinations: a CARA (Constant Absolute Risk Aversion) utility function, for which optimal allocations are not bounded, combined with a Normal distribution (centred on the optimal) of the actual allocations, and a CRRA utility function also combined with a Normal distribution centred on the optimal allocations. As we noted above, the empirical results suggest that these combinations are empirically inferior to the combination that we report; detailed results are available on request.

**Table 1** Average fitted log-likelihoods (standard deviation is in parentheses)

| | Uncorrected log-likelihoods | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | SEU | CEU | AEU | VEU | COM | Obs. |
| All | −157.29 | −154.68 | −155.82 | −154.20 | −156.73 | 129 |
| | (58.31) | (57.49) | (57.23) | (57.44) | (57.70) | |
| Treatment 1 | −159.55 | −157.48 | −158.29 | −156.83 | −159.26 | 66 |
| | (50.55) | (49.68) | (49.37) | (49.86) | (49.74) | |
| Treatment 2 | −154.93 | −151.73 | −153.23 | −151.45 | −154.08 | 63 |
| | (65.80) | (64.95) | (64.75) | (64.74) | (65.31) | |
| t-stat diff. | −0.448 | −0.566 | −0.501 | −0.531 | −0.509 | |

## 5 Results

We estimated each of the 5 preference functionals for each of the 129 subjects on a subset of the data—namely a randomly chosen 60 of the 76 problems[16]—using the constrained maximum likelihood procedure in GAUSS. We thus have, for each preference functional, for each subject, estimates of the parameters of the functional, of $s$ (the precision) and of $r$ (the risk aversion parameter). In addition, we have the maximised log-likelihood.[17] We then used, for each subject and each preference the estimated parameters to predict behaviour on the remaining 16 problems. This gives us a prediction log-likelihood for each functional and for each subject—this is, of course, a measure of the predictive ability of the theory. In the tables that follow we break down some of the summary information by Treatment; recall that Treatment 1 (66 subjects) was the less ambiguous treatment while Treatment 2 (63 subjects) was the more ambiguous treatment.

We start with Table 1 which shows the mean and standard deviation (across all subjects in each Treatment and in both Treatments) of the fitted log-likelihoods. However this table does not allow us to compare the goodness of fit *across* preference functionals, for the simple reason that they have different degrees of freedom (SEU has 4 estimated parameters, CEU 8, AEU 6, VEU 5 and COM 6). If we correct the fitted log-likelihoods for the degrees of freedom by calculating the Bayesian Information Criterion (BIC),[18] we get Table 2; recall that the *lower* the BIC the better. It seems

---

[16]Because the subjects received the 76 questions in different orders (and with the colours on the left and the colours on the right randomly selected) this means that the position of the 60 estimation questions (and hence the 16 prediction questions) varied from subject to subject, but for each subject they were randomly positioned.

[17]Note that with our combination since the variables to be explained are the *proportions* of the endowment allocated the various colours, in order to make the log-likelihoods comparable with those from other specifications, we need to subtract from the maximised log-likelihoods the sum of the natural logarithms of the amounts to be allocated in the relevant problems.

[18]This is given by $k \ln(n) - 2LL$, where $k$ is the number of estimated parameters, $n$ the number of observations and $LL$ the maximised log-likelihood.

**Table 2**  Bayesian Information Criterion (standard deviation is in parentheses)

|  | Bayesian Information Criterion | | | | | |
|  | SEU | CEU | AEU | VEU | COM | Obs. |
|---|---|---|---|---|---|---|
| All | 330.97 | 342.11 | 336.21 | 328.88 | 338.03 | 129 |
|  | (116.62) | (114.99) | (114.45) | (114.89) | (115.40) |  |
| Treatment 1 | 335.47 | 347.72 | 341.15 | 334.14 | 343.09 | 66 |
|  | (101.10) | (99.37) | (98.74) | (99.72) | (99.48) |  |
| Treatment 2 | 326.24 | 336.22 | 331.02 | 323.37 | 332.72 | 63 |
|  | (131.60) | (129.91) | (129.50) | (129.48) | (130.62) |  |
| t-stat diff. | 0.448 | 0.566 | 0.501 | 0.531 | 0.509 |  |

that, on average, VEU is the best, followed by SEU and then AEU, COM and CEU. Note that CEU is penalised by its large number of parameters.

In order to demonstrate that the means shown in Tables 1 and 2 hide rather large variations across subjects, we present histograms of the BIC across subjects in Fig. 1. We note that the shapes of the distributions of the subjects across preference functionals are very similar, and there are high correlations between the log-likelihoods across subjects over preference functionals.
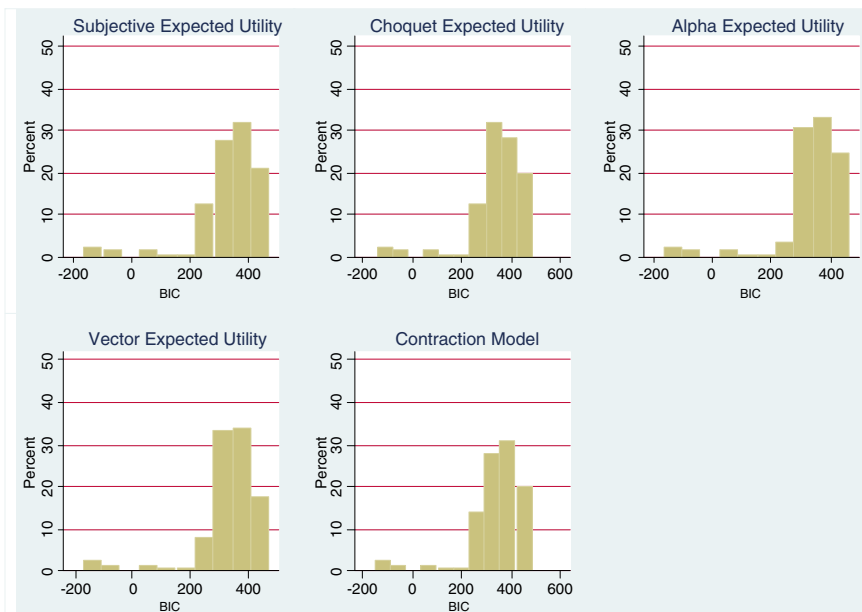


**Fig. 1**  Histograms BIC by preference functionals

**Table 3** Rankings based on the Bayesian Information Criterion (all values represent cumulative percentages - last column shows the average ranking)

| Model | 1st | 1–2 | 1–3 | 1–4 | 1–5 | Aver. Rank |
|---|---|---|---|---|---|---|
| Treatment 1 | | | | | | |
| SEU | 52 | 61 | 92 | 100 | 100 | 1.95 |
| CEU | 0 | 6 | 23 | 33 | 100 | 4.38 |
| AEU | 6 | 41 | 68 | 92 | 100 | 2.92 |
| VEU | 38 | 77 | 88 | 94 | 100 | 2.03 |
| COM | 5 | 15 | 29 | 80 | 100 | 3.71 |
| Treatment 2 | | | | | | |
| SEU | 49 | 67 | 95 | 100 | 100 | 1.89 |
| CEU | 2 | 6 | 17 | 29 | 100 | 4.46 |
| AEU | 8 | 33 | 70 | 92 | 100 | 2.97 |
| VEU | 37 | 83 | 89 | 89 | 100 | 2.03 |
| COM | 5 | 11 | 29 | 90 | 100 | 3.65 |

Tables 1 and 2 relate to averages. We now look at individual subjects. If we rank the various preference functionals using the BIC, we get Table 3. Here we report the cumulative percentage in each ranking position: so for example, in Treatment 1, SEU is ranked first for 52% of our subjects; is ranked first or second for 61% of our subjects; and so on. It is clear from this table that SEU and VEU are the 'best', then comes AEU and finally COM and CEU.

We now ask about *statistical significance* of our estimates. Because of the relationships between the preference functionals, we need to carry out two kinds of tests: nested tests and non-nested tests. We note that SEU is nested within all the other four preference functionals, but none of them (in the way that we have implemented

**Table 4** Significance test for superiority of preference functionals - five percent level

| Model | SEU | CEU | AEU | VEU | COM |
|---|---|---|---|---|---|
| Treatment 1 | | | | | |
| CEU | 26 | n.a. | 6 | 3 | 6 |
| AEU | 38 | 0 | n.a. | 2 | 3 |
| VEU | 38 | 0 | 6 | n.a. | 3 |
| COM | 14 | 0 | 9 | 3 | n.a. |
| Treatment 2 | | | | | |
| CEU | 21 | n.a. | 6 | 2 | 5 |
| AEU | 38 | 0 | n.a. | 2 | 3 |
| VEU | 38 | 0 | 13 | n.a. | 5 |
| COM | 13 | 0 | 6 | 2 | n.a. |

**Table 5** Significance test for superiority of preference functionals - one percent level

| Model | SEU | CEU | AEU | VEU | COM |
|---|---|---|---|---|---|
| Treatment 1 | | | | | |
| CEU | 14 | n.a. | 6 | 2 | 6 |
| AEU | 36 | 0 | n.a. | 2 | 3 |
| VEU | 30 | 0 | 5 | n.a. | 3 |
| COM | 11 | 0 | 5 | 3 | n.a. |
| Treatment 2 | | | | | |
| CEU | 13 | n.a. | 5 | 2 | 3 |
| AEU | 32 | 0 | n.a. | 2 | 3 |
| VEU | 33 | 0 | 6 | n.a | 3 |
| COM | 10 | 0 | 5 | 2 | n.a |

them here) are nested inside any of the others. Hence for each of CEU, AEU, VEU and COM relative to SEU a likelihood ratio test is appropriate; for each of CEU, AEU, VEU and COM against the others a Clarke test is appropriate (Clarke 2007). The results are reported in Table 4 (5% significance) and Table 5 (1% significance), where the entries are the percentage of the subjects in the two treatments (66 and 63 in total respectively). Looking at the first column of Table 4, we note that CEU and COM do relatively poorly. Indeed, they are significantly better than SEU only for 26% and 14% of the subjects, respectively, in Treatment 1 and only for 21% and 13% of the subjects, respectively, in Treatment 2 at the 5% level. Table 5 shows the

**Table 6** Average prediction log-likelihoods (standard deviation is in parentheses)

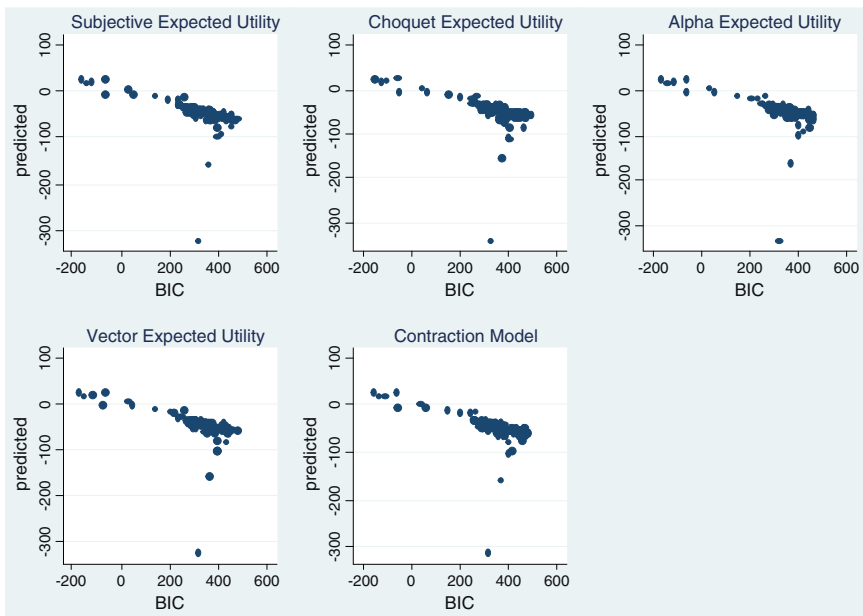| | Prediction Log-Likelihoods | | | | | |
|---|---|---|---|---|---|---|
| | SEU | CEU | AEU | VEU | COM | Obs. |
| All | −43.97 | −45.03 | −43.98 | −43.66 | −43.99 | 127 |
| | (18.57) | (19.97) | (18.43) | (18.77) | (18.66) | |
| Treatment 1 | −44.77 | −45.61 | −44.92 | −44.63 | −44.81 | 65 |
| | (15.67) | (16.89) | (15.49) | (15.67) | (15.65) | |
| Treatment 2 | −43.14 | −44.43 | −42.99 | −42.64 | −43.13 | 62 |
| | (21.29) | (22.88) | (21.16) | (21.63) | (21.48) | |
| t-stat diff. | −0.493 | −0.329 | −0.584 | −0.595 | −0.505 | |

We excluded here from the analysis two outliers (namely subjects number 14 and 22) because their inclusion significantly affects the calculation of the average and of the standard deviation, and gives a distorted picture of the results

**Table 7** Ranking based on prediction log-likelihoods (all values represent cumulative percentage - last column shows the average ranking)

| Model | 1st | 1–2 | 1–3 | 1–4 | 1–5 | Aver. Rank |
|---|---|---|---|---|---|---|
| Treatment 1 | | | | | | |
| SEU | 15 | 40 | 51 | 89 | 100 | 3.00 |
| CEU | 23 | 32 | 49 | 58 | 100 | 3.32 |
| AEU | 9 | 49 | 71 | 82 | 100 | 2.85 |
| VEU | 34 | 49 | 63 | 82 | 100 | 2.68 |
| COM | 18 | 29 | 66 | 89 | 100 | 2.92 |
| Treatment 2 | | | | | | |
| SEU | 11 | 29 | 56 | 89 | 100 | 3.10 |
| CEU | 18 | 34 | 40 | 45 | 100 | 3.57 |
| AEU | 11 | 47 | 73 | 89 | 100 | 2.76 |
| VEU | 48 | 58 | 71 | 94 | 100 | 2.25 |
| COM | 11 | 32 | 60 | 84 | 100 | 3.08 |

We excluded here from the analysis two outliers (namely subjects number 14 and 22) because their inclusion significantly affects the calculation of the average and of the standard deviation, and gives a distorted picture of the results

same results at 1% level of significance. These statistical tests on the fitting of the various preference functionals tell us that the best seem to be AEU and VEU. Indeed,
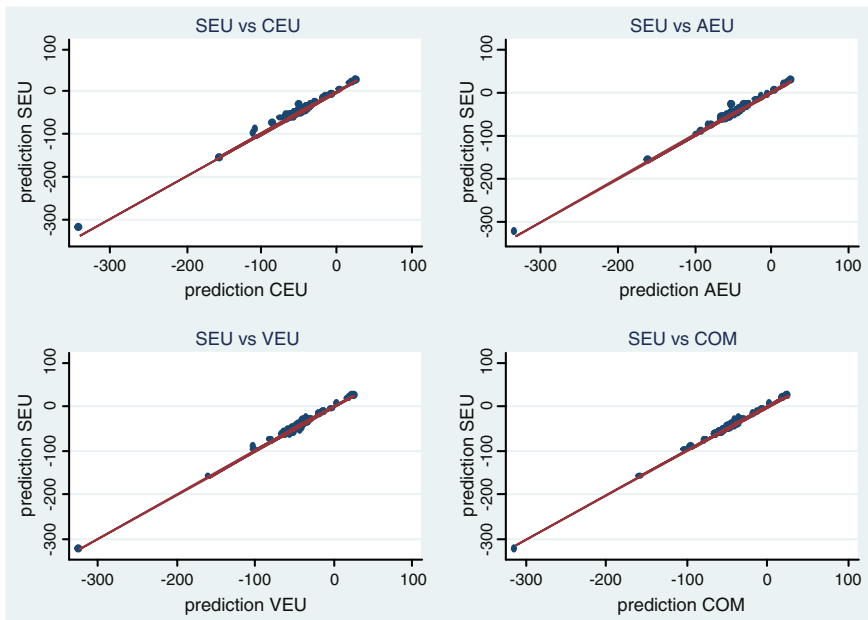


**Fig. 2** Scatter prediction log-likelihoods vs BICs

**Table 8**  Correlation between prediction log-likelihoods and BICs

| Prediction Log-Likelihoods | Bayesian Information Criterion | | | | |
|---|---|---|---|---|---|
|  | SEU | CEU | AEU | VEU | COM |
| Treatment 1 |  |  |  |  |  |
| SEU | −0.823 | −0.806 | −0.793 | −0.807 | −0.811 |
| CEU | −0.767 | −0.770 | −0.742 | −0.759 | −0.767 |
| AEU | −0.808 | −0.795 | −0.817 | −0.797 | −0.805 |
| VEU | −0.811 | −0.802 | −0.787 | −0.808 | −0.805 |
| COM | −0.809 | −0.804 | −0.787 | −0.799 | −0.807 |
| Treatment 2 |  |  |  |  |  |
| SEU | −0.928 | −0.925 | −0.923 | −0.929 | −0.927 |
| CEU | −0.896 | −0.892 | −0.891 | −0.896 | −0.896 |
| AEU | −0.908 | −0.904 | −0.925 | −0.912 | −0.910 |
| VEU | −0.899 | −0.894 | −0.897 | −0.905 | −0.898 |
| COM | −0.921 | −0.917 | −0.918 | −0.921 | −0.921 |

We excluded here from the analysis two outliers (namely subjects number 14 and 22) because their inclusion significantly affects the calculation of the average and of the standard deviation, and gives a distorted picture of the results



**Fig. 3**  Scatter prediction log-likelihoods SEU vs other preference functionals
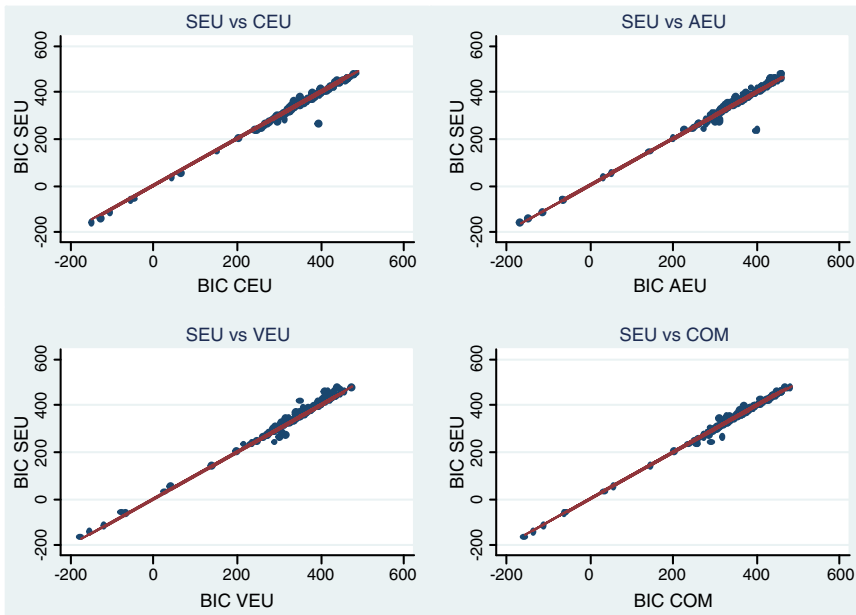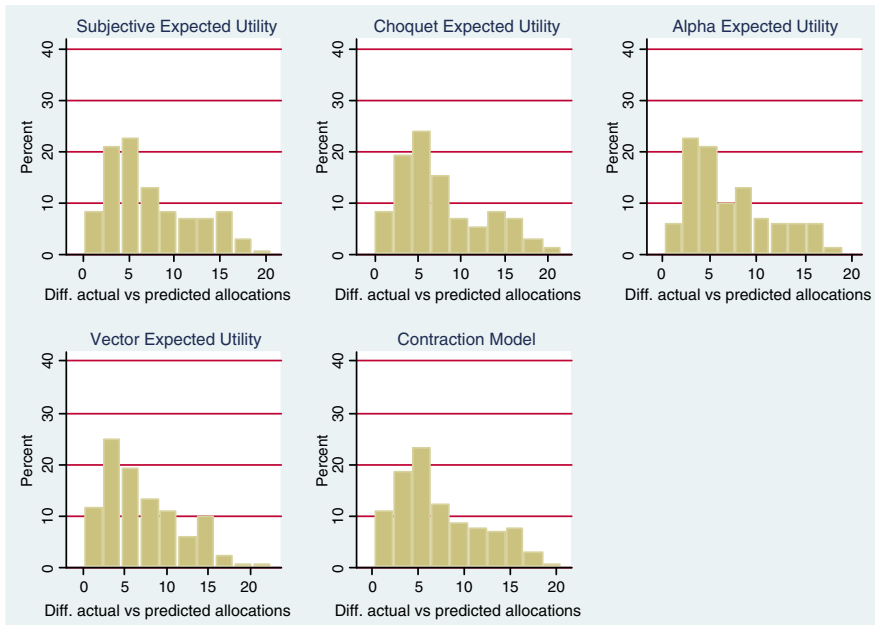
**Fig. 4** Scatter BICs SEU vs other preference functionals

these preference functionals are significantly better than SEU for 38% [36%] and 38% [30%] of the subjects, respectively, in Treatment 1, and for 38% [32%] and 38% [33%] of the subjects, respectively, in Treatment 2 at a 5% [1%] level of significance (see Table 4 [5]).

If we now turn to the prediction log-likelihoods, things are not so clear cut. From Table 6 we see that the prediction log-likelihoods have considerable variations across subjects. Table 7 gives cumulative rankings and is the counterpart of Table 3 for the fitted log-likelihoods. The findings are less clear here, with the average rankings closer together. We note that we are using *uncorrected* log-likelihoods (which seems appropriate as we are concerned here with predictions). What is particularly striking

**Table 9** Average departure from best prediction (standard deviation is in parentheses)

|             | SEU    | CEU    | AEU    | VEU    | COM    |
|-------------|--------|--------|--------|--------|--------|
| All         | 7.37   | 7.63   | 7.34   | 7.32   | 7.34   |
|             | (4.65) | (4.93) | (4.58) | (4.69) | (4.66) |
| Treatment 1 | 6.43   | 6.49   | 6.51   | 6.49   | 6.34   |
|             | (3.89) | (4.04) | (4.01) | (3.94) | (3.88) |
| Treatment 2 | 8.36   | 8.82   | 8.20   | 8.18   | 8.39   |
|             | (5.18) | (5.49) | (4.99) | (5.25) | (5.19) |

**Fig. 5** Differences actual vs predicted allocations

is that, while VEU still seems to be particularly good with its ranking better than the rankings of the others, the average rankings are much closer together than they were for the BICs. At the same time, there is a very strong correlation between the BICs (corrected log-likelihoods) and the prediction log-likelihoods, as Fig. 2 makes clear: this suggests that the best (BIC) fitting preference functional is often also the best predicting log-likelihood, as Table 8 confirms. At the same time, Fig. 3 emphasises that there is a very high correlation over preference functionals (for any one subject) and hence there are very small differences between the goodness of prediction of the various preference functionals. But this was also the case for the Bayesian Information Criteria: there is much more variation across subjects than across preference functionals (Fig. 4). This latter point is emphasised by Table 9 and Fig. 5. Table 9 presents the average prediction error (as measured, for any one subject and preference functional, by the square root of the mean squared difference between the actual allocation and the predicted allocation using that preference functional). Figure 5 presents the distribution across subjects of this measure of prediction error. What is noticeable is the big difference across subjects and the small difference across preference functionals.

We conclude that in terms of predictions, VEU and (to a lesser extent) AEU seem to be better than SEU though, in terms of magnitudes, they are not a lot better than SEU. The loss in predictive power in using SEU is relatively small in magnitude.

## 6 Conclusions

We conclude that VEU and, to a lesser extent, AEU are somewhat better than SEU. However what appears to be rather odd is that there does not seem to be a treatment effect with respect to preference functionals: relatively the different preference

**Table 10** Descriptive statistics average estimated parameters

| Theory | Parameter | All | Treatment 1 | Treatment 2 | t-test diff. |
|---|---|---|---|---|---|
| SEU | $p_1$ | 0.231 | 0.223 | 0.239 | −0.989 |
| | $p_2$ | 0.331 | 0.317 | 0.345 | −2.912 |
| | $p_3$ | 0.439 | 0.460 | 0.416 | 2.385 |
| | $r$ | 2.223 | 2.129 | 2.322 | −0.229 |
| | $s$ | 20.543 | 21.978 | 19.040 | 0.815 |
| CEU | $w_{E_1}$ | 0.222 | 0.210 | 0.234 | −1.241 |
| | $w_{E_2}$ | 0.301 | 0.290 | 0.313 | −1.448 |
| | $w_{E_3}$ | 0.408 | 0.435 | 0.380 | 2.639 |
| | $w_{E_2 \cup E_3}$ | 0.759 | 0.770 | 0.747 | 0.911 |
| | $w_{E_1 \cup E_3}$ | 0.613 | 0.635 | 0.589 | 2.490 |
| | $w_{E_1 \cup E_2}$ | 0.542 | 0.489 | 0.597 | −3.695 |
| | $r$ | 1.999 | 1.969 | 2.031 | −0.072 |
| | $s$ | 20.897 | 22.020 | 19.721 | 0.649 |
| AEU | $\underline{p}_1$ | 0.204 | 0.194 | 0.214 | −1.315 |
| | $\underline{p}_2$ | 0.300 | 0.286 | 0.315 | −2.465 |
| | $\underline{p}_3$ | 0.404 | 0.424 | 0.383 | 1.836 |
| | $\alpha$ | 0.344 | 0.399 | 0.286 | 1.907 |
| | $r$ | 2.201 | 2.088 | 2.320 | −0.281 |
| | $s$ | 19.359 | 20.281 | 18.393 | 0.575 |
| VEU | $p_1$ | 0.233 | 0.224 | 0.241 | −1.036 |
| | $p_2$ | 0.330 | 0.317 | 0.343 | −2.546 |
| | $p_3$ | 0.438 | 0.459 | 0.416 | 2.291 |
| | $\delta$ | −0.020 | −0.014 | −0.027 | 1.130 |
| | $r$ | 2.085 | 2.027 | 2.145 | −0.156 |
| | $s$ | 20.251 | 21.408 | 19.038 | 0.687 |
| COM | $\underline{p}_1$ | 0.203 | 0.192 | 0.213 | −1.254 |
| | $\underline{p}_2$ | 0.300 | 0.284 | 0.316 | −2.328 |
| | $\underline{p}_3$ | 0.408 | 0.427 | 0.388 | 1.675 |
| | $\lambda$ | 0.456 | 0.474 | 0.436 | 0.826 |
| | $r$ | 1.961 | 1.854 | 2.073 | −0.345 |
| | $s$ | 19.666 | 20.687 | 18.596 | 0.625 |

functionals perform similarly in the two treatments. In light of the nature of the results, we can legitimately ask: is there no treatment effect?

In order to explore this question, we present Table 10. This gives the averages of the estimated parameters for each preference functional for each treatment. Let us look at SEU, as a similar message emerges (appropriately modified) for the other functionals.[19] The mean estimates of risk aversion are not significantly different in the two treatments; neither are the mean estimates of the noise (though the noise is slightly higher in Treatment 2). But examine the estimated probabilities: in Treatment 1 the mean estimates of the three probabilities (pink, yellow and blue) are 0.223, 0.317 and 0.460; in Treatment 2, 0.239, 0.345 and 0.416. The true probabilities are 0.2, 0.3 and 0.5. There are significant differences between the mean estimates in Treatment 1 and those in Treatment 2. In Treatment 1 the mean estimated probabilities are close to the true ones. In Treatment 2 they are significantly closer to equal probabilities for the three colours. *Subjects were responding to the ambiguity by working on the basis of almost equal probabilities.* So they were not working with a more sophisticated preference functional in the more complicated environment of Treatment 2. On the contrary, they responded by simplifying their decision problem.

A clear possible criticism of our experiment relates to the question of whether or not our subjects were ambiguity-neutral. If they were, then it rather trivially follows that SEU has no worse chance of explaining behaviour than the other preference functionals. Take CEU for example: if a subject were ambiguity-neutral then the estimated capacities would be additive. There are various ways that one can test this (like testing whether the capacities depart significantly from additivity), but the simplest is to test whether the CEU model fits significantly better than SEU. Table 4 and 5 give the answer: for 26% (14%) of the subjects in Treatment 1 CEU fits significantly better than SEU at the 5% (1%) level and for 21% (13%) in Treatment 2. But this is a relatively weak test because of the large number of parameters in CEU. A much clearer picture emerges from VEU and AEU. For VEU the corresponding figures are 38% (30%) in Treatment 1 and 38% (33%) in Treatment 2. For AEU the corresponding figures are 38% (36%) in Treatment 1 and 38% (32%) in Treatment 2. So we can conclude that many of our subjects were not ambiguity-neutral.

Another possible criticism relates to whether the experiment was designed in such a way so that it was unable to distinguish between the different preference functionals. It is clear that the preference functionals are different (and thus lead to different

---

[19]With the other preference functionals we note the following, as far as the mean parameters are concerned:

(1)   with CEU the estimated mean capacities are almost additive, but get slightly less so in Treatment 2;

(2)   with AEU, the mean lower bounds on the probabilities are close to the SEU subjective probabilities, and get closer to equality in Treatment 2;

(3)   with VEU the mean $\delta$ parameter is close to 0 and similar in the two treatments;

(4)   with COM the mean lower bounds on the probabilities are close to the SEU probabilities and the $\lambda$ is close to 0.5;

(5)   we note that the $\alpha$ parameter in the AEU model is on average lower than the $\lambda$ parameter in the COM model. This suggests that the Steiner point is a less important consideration to the subjects than the maximum expected utility.

**Table 11** Bayesian Information Criterion when the row model is the true one

| True Model | Bayesian Information Criterion | | | | |
| | SEU | CEU | AEU | VEU | COM |
| --- | --- | --- | --- | --- | --- |
| SEU | 471.21 | 484.64 | 479.03 | 474.46 | 479.40 |
| CEU | 529.03 | 440.82 | 537.69 | 449.09 | 530.93 |
| AEU | 561.17 | 496.39 | 469.24 | 514.19 | 569.83 |
| VEU | 487.69 | 456.30 | 496.35 | 446.31 | 478.52 |
| COM | 485.44 | 471.19 | 493.40 | 475.53 | 465.80 |

decisions), which implies that the models are distinguishable given our parameterisations: we are simply reinforcing here the theoretical point we have made earlier, at the end of Section 1. We return to this point since the theory ignores the existence of noise in subjects' decisions. However there is noise in subjects' decisions and this noise could drown out the distinguishability that clearly exists in the absence of noise, unless the number of problems posed to the subjects was sufficiently high and the problems appropriately constructed. So before we implemented the experiment we carried out a simulation assuming a particular (realistic) level of noise and the problems that we asked. With our parameterisation of the preference functionals we can demonstrate with a simple example, shown in Table 11, that the preference functionals under analysis are fully distinguishable. In this example we have assumed a reasonable set of parameters for each functional, a particular specification and a realistic value for the precision $s$[20], and simulated estimation with 100 repetitions. Each cell reports the mean Bayesian Information Criterion (the lower the better) of the column model when the row model is the true one. Table 11 shows that the preference functionals under analysis are fully distinguishable. Indeed, in each row the diagonal element is always a lot smaller than the off-diagonal elements. This means, for example, that if we know that a specific subject has CEU preferences, then CEU best fits behaviour in the experiment. Obviously this is for subjects who are 'clearly' CEU: if a subject has CEU preferences that are 'close' to SEU preferences and there is a lot of noise in that subject's behaviour, then distinguishability is more problematic. But, of course, if that is the case then SEU predictions will also be 'close' to CEU predictions.

This property of the experimental design (distinguishability) was not simply by chance as we carried out intensive pre-experimental simulations in order to select the set of problems to ask the subjects. The purpose of these simulations was precisely to select a number and a set of problems which would enable us to discriminate between the preference functionals, given the amount of noise in the subjects' responses.

[20]We run 100 replications using a coefficient of risk aversion r equal to 0.8 and a coefficient of precision equal to 12. For each preference functional we set the following parameters' values: SEU: $p_1 = 0.2$, $p_2 = 0.3$, $p_3 = 0.5$; CEU: $w_{E(1)} = 0.10$, $w_{E(2)} = 0.20$, $w_{E(3)} = 0.30$, $w_{E(2) \cup E(3)} = 0.85$, $w_{E(3) \cup E(1)} = 0.75$, $wE(2) \cup E(2) = 0.65$; AEU: $\underline{p}_1 = 0.10$, $\underline{p}_2 = 0.15$, $\underline{p}_3 = 0.25$, $\alpha = 0.5$; VEU: $p_1 = 0.2$, $p_2 = 0.3$, $p_3 = 0.5$, $\delta = 0.10$; COM: $\underline{p}_1 = 0.10$, $\underline{p}_2 = 0.20$, $\underline{p}_3 = 0.30$, $\lambda = 0.75$.

Clearly the greater the noise the more problems are required; this explains the relatively large number of problems in our experiment. We had carried out a pilot experiment to determine how much noise there was in behaviour, and this informed the simulation and hence the number and choice of problems.

Before we conclude we should note that the fitting and prediction parts of the exercise give somewhat different results. In terms of fitting, one of the more general models (VEU) does seem to fit better than SEU for many of the subjects. However, when it comes to prediction, as is clearly shown in Table 9, some of this superiority disappears (though AEU and VEU are still marginally better overall than SEU). Hence one might not lose a lot in predictive power in using SEU rather than one of the more general functionals. An econometrician might regard this as an inevitable consequence of over-fitting: as is well-known, if one fits an $n^{\text{th}}$-degree polynomial to n observations from a truly linear relationship (with noise), the fit is better than a linear fit, but extrapolative predictions are almost certainly worse. It should be noted, however, that our prediction problems were a randomly chosen subset of all the problems, and cannot be considered as extrapolative. Indeed it is not clear in our case what "extrapolative problems" means.

So the bottom line appears to be that VEU and AEU are better than SEU in terms of explanation (less so COM and CEU), and that some, but not all, of this superiority disappears when it comes to prediction. Moreover, when we move from Treatment 1 (almost a case of risk) to Treatment 2 (clearly a situation of ambiguity) subjects do not respond by moving to a more sophisticated preference functional. Instead they seem to respond by having subjective probabilities further away from the true probabilities and nearer to equality. In our view, this is a rational response: if a situation is ambiguous, and hence complicated, why complicate it further by using a more complicated preference functional? In this respect we note that the best functional, VEU, in the way that we have parameterised it, is a particularly simple extension of SEU. If we look at Eq. 12 above we see that it is essentially SEU with the probability attached to the best outcome decreased by a small amount $\delta$ and the probability of the worst outcome increased by a small amount $\delta$, this $\delta$ depending upon the amount of ambiguity and the subject's perception and reaction to it. Of course, this $\delta$ could be negative if a person is ambiguity-loving. Indeed for 49 of our subjects the estimated value of $\delta$ was positive and for 80 it was negative. On average, as we see from Table 10, this $\delta$ increased slightly in absolute value. At the same time, the average values of the estimated probabilities moved closer to equality. This seems a rather sensible way to respond to increased ambiguity.

## References

Abdellaoui, M., Baillon, A., Placido, L., Wakker, P. (2011). The rich domain of uncertainty: source functions and their experimental implementation. *American Economic Review*, *101*, 695–723.

Ahn, D.S., Choi, S., Gale, D., Kariv, S. (2010). Estimating ambiguity aversion in a portfolio choice experiment. Working Paper.

Andersen, S., Fountain, J., Harrison, G.W., Rutström, E.E. (2009). Estimating aversion to uncertainty, Working Paper.

Andersen, S., Harrison, G.W., Lau, M.I., Rutstrom, E.E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, *9*, 383–405.

Andreoni, J., & Miller, J. (2002). Giving according to garp: an experimental test of the consistency of preferences for altruism. *Econometrica*, *70*(2), 737–753.

Camerer, C. (1995). Individual decision making. In J. Kagel & A. Roth (Eds.) *Handbook of experimental economics* (pp. 587–703). Princeton University Press.

Camerer, C., & Weber, M. (1992). Recent development in modeling preferences: uncertainty and ambiguity. *Journal of Risk and Uncertainty*, *5*(4), 325–370.

Chambers, R.G., Melkonyan, T., Pick, D. (2010). Experimental evidence on multiple prior models in the presence of uncertainty, Working Paper.

Choi, S., Fisman, R., Gale, D., Kariv, S. (2007). Consistency and heterogeneity of individual behavior under uncertainty. *American Economic Review*, *97*(4), 1921–1938.

Clarke, K.A. (2007). A simple distribution-free test for non-nested model selection. *Political Analysis*, *15*, 347–363.

Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms. *Quarterly Journal of Economics*, *75*, 643–669.

Etner, J., Jeleva, M., Tallon, J.M. (2012). Decision theory under ambiguity. *Journal of Economic Surveys*, *26*(2), 234–270.

Fox, C.R., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *Quarterly Journal of Economics*, *110*(3), 585–603.

Gajdos, T., Hayashi, T., Tallon, J.M., Vergnaud, J.C. (2008). Attitude toward imprecise information. *Journal of Economic Theory*, *140*, 27–65.

Ghirardato, P., Maccheroni, F., Marinacci, M. (2004). Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory*, *118*(2), 133–173.

Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, *18*, 141–153.

Greiner, B. (2004). The online recruitment system orsee 2.0 – a guide for the organization of experiments in economics. University of Cologne Discussion Paper. www.orsee.org.

Halevy, Y. (2007). Ellsberg revisited: an experimental study. *Econometrica*, *75*(2), 503–536.

Hey, J.D., Lotito, G., Maffioletti, A. (2010). The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity. *Journal of Risk and Uncertainty*, *41*(2), 81–111.

Holt, C.A., & Laury, S.K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Karni, E., & Safra, Z. (1987). Preference reversals and the observability of preferences by experimental methods. *Econometrica*, *55*, 675–685.

Loomes, G. (1991). Evidence of a new violation of the independence axiom. *Journal of Risk and Uncertainty*, *4*(1), 91–108.

Maccheroni, F., Marinacci, M., Rustichini, A. (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, *74*, 1447–1498.

Morone, A., & Ozdemir, O. (2012). Displaying uncertainty information about probability: experimental evidence. *Bulletin of Economic Research*, *64*, 157–171.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, *57*(3), 571–587.

Segal, U. (1987). The Ellsberg paradox and risk aversion: an anticipated utility approach. *International Economic Review*, *28*, 175–202.

Siniscalchi, M. (2009). Vector expected utility and attitudes toward variation. *Econometrica*, *77*(3), 801–855.

Wakker, P.P. (2010). *Prospect theory for risk and ambiguity*. Cambridge: Cambridge University Press.

Wilcox, N. (2007). *Predicting risky choices out-of-context: a monte carlo study*. University of Houston Working Paper.

Wilcox, N. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. In J.C. Cox & G.W. Harrison (Eds.)*, Research in experimental economics: risk aversion in experiments* (Vol. 12, pp. 197–292). Bingley, UK: Emerald.

Wilcox, N. (2011). Stochastically more risk averse: a contextual theory of stochastic discrete choice under risk. *Journal of Econometrics*, *162*, 89–104.