Weighting probabilities in ambiguity

Dmitri Vinogradov

Preliminary draft,
please do not quote

September 2013

Abstract

Probability weighting function (as in prospect theory) is a transformation of probabilities into decision weights to reflect the inability of subjects to reason in terms of probabilities directly. Ambiguity is often associated with unknown probabilities, hence there is aparently nothing to transform. We adopt the view of decision-making in ambiguity that assumes that subjects form [second-order] expectations about feasible probability distributions. Drawing on the data from several two-urn Ellsberg experiments, we elicit second order probabilities that govern these expectations for subjects with different ambiguity attitudes, and derive their weighting functions for second-order probabilities. These weighting functions are similar to those observed in the prospect theory with regards to primary probabilities. The typical inverse-S shape of the probability weighting function is primarily due to ambiguity aversion. Moreover, expected-utiility maximisers (ambiguity-neutral subjects) who are typically assumed to reason in probabilities, comply with the theory before they face any signals, yet exhibit an even higher degree of pessimism after the signals; their probability weighting function becomes non-linear.

JEL Classification: D81

Keywords: ambiguity, Ellsberg paradox, second-order probabilities, probability weighting

Address for Correspondence: Author's address

1 Introduction

In prospect theory (Kahneman and Tversky, 1979 and 1992), decision-makers (DM) value outcomes relative to a reference point and distort probabilities by applying a so-called probability weighting function. The current paper focuses on the distortions of probabilities. With regards to the situations of risk (Arrovian uncertainty), in which probability distributions are given, there is no obvious reason not to rely on the "true" probabilities, hence deviations from them seem irrational. Yet it is not perfectly clear what is the source of this irrationality. We place a DM in an ambiguous environment (Knightian uncertainty, Knight, 1921) and employ the "second-order" approach to decisions in ambiguity (SOEU) independently and differently axiomatized in Klibanoff, Mukerji and Marinacci (2005), Nau (2006), Chew and Sagi (2008), and Ergin and Gul (2009). We extend this approach to encompass probability weighting in ambiguity.

The original (non-cumulative) prospect theory is able to accommodate Ellsberg paradox through the probability weighting function, yet the typical functional form implies ambiguity-seeking behaviour, on top of the usual critique of the violation of stochastic dominance. The cumulative prospect theory (and generally rank-dependent decision theory) fails to accommodate the Ellsberg paradox unless one replaces the weighting function with subadditive probability weights (Wakker, 2010, provides an extensive analysis of the applicability of the prospect theory to ambiguity, see also Appendix A) However the latter contradicts most empirical studies of the prospect theory that suggest that the probability weighting is a well defined inverse-S shaped function, and since it maps probabilities into probabilities, it is additive, unlike capacities. The extended SOEU resolves this paradox. Prospect theory obtains as a special case, in which the DM is not certain about the probability distribution even if the latter is given. This yields useful insights in the nature of probability weighting functions. In particular, we show that probability weighting reflects subjects' ambiguity attitude, and hence prospect theory implicitly considers subjects that create "psychological ambiguiy". These implications of the extended SOEU are confirmed through a series of experiments thus lending empirical support to the second order approach.

Intuitively, a subject with EU-preferences, has probability weights identically equal to probabilities themselves. Any deviations from this are only due to the existence of subjects with non-EU preferences. Imagine now the same EU subject facing ambiguity and trying to reason in terms of secondary probabilities. First, this is possible: if decision-makers can form subjective probabilities about the possible relevant probability distributions, then subjective probabilities of an EU-decision-maker are a special case. Second, similarly to the prospect theory, any deviations from the decisions of this EU-subject should be due to non-EU-reasoning, as in the prospect theory example above. The only difference between the two groups of subjects are subjective probabilities. Subjective probabilities of non-EU subjects are related to the subjective probabilities of the EU-subject by a probability weighting function. This idea outlines the main approach of the current study.

We conduct a series of standard two-urn Ellsberg experiments (urn A contains 100 balls of Red and Black colours, urn B contains 50 Red balls and 50 Black balls, subjects obtain a prize if the ball drawn from their selected urn is of the designated colour). Subjects' second-order (subjective) probability distributions about the distribution of the balls in the ambiguous urn (urn A) are elicited using an approach similar to Carlson-Parkin (1975) method for quantifying inflation expectations from qualitative data. The elicited distribution for ambiguity-neutral (EU-consistent) subjects, as expected, has a mean close to .5 and is taken as a benchmark. The mean of the elicited distribution for the whole sample lies below .5, which explains ambiguity aversion on average. The transformation of the former into the latter is described by an inverse S-shaped function with the standard properties as in the prospect theory. This suggests that non-EU subjects treat second-order probabilities in the same way as described in the prospect theory for primary probabilities. Fig 1 reproduces typical shapes of probability weighting functions from fig. 1. in Weber (1994).[1] Our result for the weighting functions of second-order probabilities confirms that ambiguity attitude is reflected in the weighting function.

[1] The labels "optimistic" and "pessimistic" are reversed in the original figure. This mistake is confirmed with the author in a private correspondence, and corrected here.

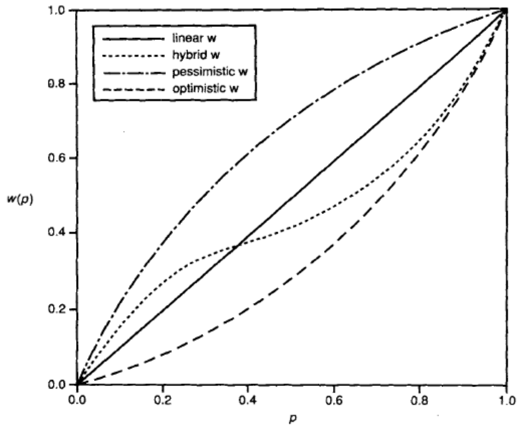

Figure 1. Optimistic, pessimistic, hybrid, and linear probability weighting functions w used in the mapping of cumulative probabilities p into rank-dependent decision weights $w(p)$. Reproduced from Weber (1994), fig. 1, p.233, with correct labels for pessimistic and optimistic weighting functions.

By the design of the experiment, we are able to distinguish between ambiguity-averse and ambiguity-loving subjects from the whole sample, and consequently derive the probability weighting functions for both groups. The standard inverse-S shape only holds for ambiguity averse subjects (pessimists), whereas for ambiguity-lovers (optimists) the weighting function is direct-S shaped. These effects of optimism and pessimism on the weighting function were first predicted by Quiggin (1982) who notes "pessimism is rather difficult to distinguis empirically from risk-aversion" (p.335), which is because Quiggin defines pessimism (optimism) as ascribing greater weights to worse (better) outcomes, i.e. outcomes with lower (higher) ranks.[2] The application of the weighting function to second order probabilities overcomes this difficulty, lending the usual for the ambiguity literature interpretation of optimism and pessimism (as above).

Furthermore, the we study how the arrival of new information impacts subjects' probability weighting functions. It turns out that the arrival of news is best reflected in the changes of the probability weighting function, not directly in the assumed probabilities that enter the decision functional. For ambiguity neutral subjects this suggests

[2] Weber (1994) calls the same effects within the prospect theory rather "extremity weighting" to replace "optimism" (because it shifts the weight towards larger gains or losses) and "status quo weighting" to replace "pessimism" (because it gives a higher weight to.outcomes that only minorly change the status quo, i.e. small gains and losses).

an interesting application of NEO-additive capacities (Chateauneuf et al., 2007, 2009) which can be seen as a weighting function that preserves linearity in probabilities, yet allows for overweighting of low and underweighting of high probabilities.

2 Second-order probabilities

Consider a "smooth" decisionmaker, as in Klibanoff et al. (2005) with a decision functional of the form

$$V(f) = \int_{\Delta} \phi \left(\int_S u(f)\, d\pi \right) d\mu = \mathbf{E}_{\mu} \phi \left(\mathbf{E}_{\pi} u \circ f \right), \tag{1}$$

where f is a Savage act defined on a state space S, which maps S on the set of consequences C, u is a vN-M utility function, π is a probability measure on S, ϕ is a map from reals to reals, μ is the decision-maker's subjective prior over Δ, the possible probabilities over S. Models by Nau (2006), Chew and Sagi (2008), and Ergin and Gul (2009) and other are derived under different conditions but in the context of the current paper deliver essentially the same functional form, representing the expected (value of the) expected utility. Note that in Nau (2006) the latter is a special case, whereas the general model does not require separability of second-order probabilities from the value function.

2.1 Special cases: probability weighting

As a generalization, Klibanoff et al. (2005) suggest

$$V(f) = \int_{\Delta} \phi \left(\int_S u(f)\, d\psi(\pi) \right) d\mu, \tag{2}$$

to capture a probability distortion $\psi : [0,1] \to [0,1]$, as if preferences over lotteries were RDU preferences. This generalisation is possible if their Axiom 1 (existence of an expected utility on lotteries) is relaxed.

Consider another generalization, which is obtainable with a relevant amendment of Axiom 2 (Subjective expected utility on 2nd order acts) to allow for distorted beliefs $\psi(\mu)$, where ψ is a mapping as above:

$$V(f) = \int_{\Delta} \phi \left(\int_S u(f)\, d\pi \right) d\psi(\mu). \tag{3}$$

Here the decision-maker has EU-preferences over lotteries, yet it is the set of "subjective probabilities" that is distorted. This makes sense for the following reason: for the decision-maker, probabilities π are the "figments of imagination". When making a decision in terms of second-order acts (hence second-order probabilities and utilities), a decision-maker has to *imagine* a probability distribution π on S, and then decide *how possible* this particular probability distribution is. Why should someone apply a distortion ψ to particular values of $\pi(s)$ that have been *imagined*? Why not assuming a "distorted" value from the beginning if it should be "distorted" (as compared to what?) at all? If a decision-maker imagines a particular probability distribution, there is no way (s)he can *act* as if a different distribution was imagined: (1) the act is governed by second-order probabilities, and (2) if this probability distribution is unthinkable from the perspective of the decision-maker, and hence needs to be distorted in any way to become "more realistic" from his/her perspective, (s)he would assign a probability of zero to this probability and correspondingly increase the probability of the distribution that corresponds to the "realistic" distortion of this one: there exists μ' such that

$$V(f) = \int_{\Delta} \phi \left(\int_S u(f)\, d\psi(\pi) \right) d\mu(\pi) = \tag{4}$$

$$= \int_{\Delta} \phi \left(\int_S u(f)\, d\psi \right) d\mu'(\psi) \equiv \int_{\Delta} \phi \left(\int_S u(f)\, d\pi \right) d\mu', \tag{5}$$

and hence imposing a distortion "inside" the functional does not really generalize (1).

Instead, (3) is a generalization of (1): the second part of the decision-making process relates to the decision on the "possibility" of each particular distribution. This can be induced by external signals or other sources of information. It is logical to assume that a decision-maker, when receiving a probabilistic signal of second-order, perceives it in a distorted way, thus applying the distortion map ψ to "subjective probabilities" μ instead of "imagined probabilities" π.

2.2 Descrete case

A special case of (3) for descrete states $s = 1..S$ with consequences $c_s = f(s)$ and set Δ of probabilities μ_j $(j = 1..\Delta)$ over probability distributions $\pi_j = (\pi_j(1), .., \pi_j(S))$

can be written as

$$V\left(c_{1},..c_{S}\right)=\sum_{j=1}^{\Delta}\psi\left(\mu_{j}\right)\cdot\phi\left(\sum_{s=1}^{S}\pi_{j}\left(s\right)u\left(c_{s}\right)\right). \tag{6}$$

If a decision-maker is absolutely aware of a particular probability distribution π_j, he assigns $\psi\left(\mu_j\right)=1$, and his decisons correspond to the vN-M expected utility maximization:

$$V\left(c_{1},..c_{S}\right)=\sum_{s=1}^{S}\pi_{j}\left(s\right)u\left(c_{s}\right). \tag{7}$$

A prospect theory representation obtains if a decision-maker considers only degenerate probability distributions $\pi_j=\left(0,..,0,\underset{j}{1},0,..0\right)$ and hence each μ_j effectively detemines the probability of consequence c_s in state $s=j$, which yields

$$V\left(c_{1},..c_{S}\right)=\sum_{s=1}^{S}\psi\left(\mu_{s}\right)\cdot\phi\left(u\left(c_{s}\right)\right). \tag{8}$$

Here the probabilities are distorted by an [equivalent of the] probability weighting function as in the prospect theory. Since both EU and PT versions are derived from a special case of SOEU, the provides an interpretation of why and when a decision-maker exhibits a prospect-theory type of misperception of probabilities. When a decision-maker faces no uncertainty (as in 7) he behaves as a vN-M EU-maximiser. According to (8), a decision-maker can only employ a probability-weighting function if he views the outcomes as lotteries (limit cases with $\varepsilon\to 0$ for probability distribution $\pi_j\left(s_j\right)=1-\varepsilon$ for state s_j with outcome c_j and probability ε for all other outcomes c_i). The probability weighting is applied to second-order probabilities, instead of probabilities of states (consequences), though yielding a representation similar to the original prospect theory.

2.3 Small worlds in small groups

A typical approach in the existing models of decision-making in ambiguity is to axiomatize the behaviour of a single decision-maker to capture ambiguity attitude at an individual level. In this paper we rather consider an average decision-maker, which equips us with a parsimonious interpretation of second-order probabilities. This has parallels with the "small worlds" as used in Chew and Sagi (2008), who in turn extend the concept originally coined in by Savage (1954). They derive a general theory of decisions in ambiguity, in which decision-makers may have different subjective views on the situation of choice. In a "big world", subjects consider all possible states, whereas in a "small world" they may disregard the states that are irrelevant from a particular perspective. It is the combination of various perspectives (and resulting small worlds) that yields a two-dimensional utility representation: one dimension refers to the decision rule in each of the small worlds, and another dimension refers to the cross-section of all small worlds. The decision rule within a single small world is described by a utility function over lotteries induced by given acts (Theorem 2 in Chew and Sagi, 2008), whereas the probability measure for this lottery is uniquely determined by the small world (Theorem 1 and Definition 6 in Chew and Sagi, 2008). As a special case, this yields a two-stage representation similar to (6), yet here we suggest a different interpretation of it.

Consider a large sample of heterogeneous decision-makers, each of whom acts as if they have subjective beliefs $\pi_j(s)$, identical for all members of a particular group j but different across the groups. Since in the model of Chew and Sagi (2008) each small world generates a unique probability measure, we can see this group of decision makers as living in the same small world. Now construct an average decision-maker as follows. Denote with ψ_j the weight of group j in the whole sample and $U(L_{\pi_j})$ the utility value that group j assigns to lottery L_{π_j} generated by their small world's probability measure π_j. Decision-makers can be said to have *on average* a utility of $V = \sum_{j=1}^{\Delta} \psi_j \cdot U(L_{\pi_j})$. In the small worlds of Chew and Sagi (2008) ϕ is non-satiated in probability[3] but, contrast to the EU framework, does not need to be linear in it. A utility function $U(L_{\pi_j}) = \phi\left(\sum_{s=1}^{S} \pi_j(s)\, u(c_s)\right)$ meets this condition. Specify weights to be dependent on the mass μ_j of group j, i.e. $\psi_j = \psi(\mu_j)$ but not necessarily equal to the relative size of each group j in the whole sample. This yields a representation identical to (6).

$$V = \sum_{j=1}^{\Delta} \psi(\mu_j) \cdot \phi\left(\sum_{s=1}^{S} \pi_j(s)\, u(c_s)\right). \tag{9}$$

[3] A real valued function ϕ defined on lotteries $L_p(x, x') = (p, x; 1-p, x')$ is non-satiated in probability if, whenever $p, p+q \in (0,1)$ the equality $U(L_p(x,x')) = U(L_{p+q}(x,x'))$ for all $x, x' \in X$ implies $q = 0$.

This rule is now a *definition* of the **average decision-maker** in the sample. This decision-maker can be seen as being aware of the existence of various small worlds but uncertain about which of them he might belong to. To construct a decision rule, he therefore weights the possibility of belonging to each of the groups, based on the relative size of them, and smoothes the values of utilities achieved in any of the small worlds. Function ϕ has a smoothing effect if it it is concave, i.e. the differences between large and small values of $u\left(L_{\pi_j}\right)$ diminish once ϕ is applied, which corresponds to the usual notion risk-aversion, yet in this case applied to the uncertainty of belonging to one of the small worlds. In the spirit of the Klibanoff et all. (2005) interpretation, this corresponds to ambiguity aversion, whereas convex ϕ corresponds to ambiguity loving behaviour as it exacerbates the differences between different small worlds. Note that ambiguity attitude in this context describes the above defined average decision maker, and does not need to apply to any individual member of the sample (unless they behave similarly to the average decision-maker, i.e. know they can be of any type j, possibility of which is measured by $\psi\left(\mu_j\right)$).

A particular form of the weighting function is $\psi\left(\mu_j\right) = \psi^C\left(\sum_{k=1}^{j}\mu_k\right) - \psi^C\left(\sum_{k=1}^{j-1}\mu_k\right)$, matching the one used in the cumulative prospect theory (therefore superscript C).

By construction, the weighting transformation $\psi\left(\mu_j\right)$ is applied to the relative masses μ_j of groups j. It may still be possible to weight probability measures π_j as in (2) but this has nothing to do with the definition of the average subject, as it refers to the behaviour of decision-makers in a particular group. The interpretation of the second-order probabilities from the perspective of an average decision-maker is useful in the following experimental elicitation of probability weights.

If all small worlds yield exactly the same probability measure π_j, or all small worlds except that of group j are seen as impossible ($\psi\left(\mu_j\right) = 1$) then $EU = \phi^{-1}(V) = \sum_{s=1}^{S}\pi_j(s)\,u(c_s)$ is the decision functional.[4]

[4] This property also holds in other second-order utility models. In fact, in an unambiguos environment ϕ is irrelevant as it is responsible only for ambiguity attitude, which does not matter as long as there is no ambiguity.

3 Elicitation of probability weights

A nice feature of the probability weighting function in the prospect theory is that it captures the psychological aspects of the perception of probabilities. Herold and Netzer (2011) even suggest that weighting functions arise as a second-best optimum in an evolutionary framework. This illustrates that unlike second-order probabilities themselves, their transformation through the weighting function can be explained rather than just assumed. Still, an experimental identification of the weighting function requires the knowledge of underlying probabilities (the very subject of weighting). In traditional prospect theory experiments these are given by the experimentator, which is not the case for second-order probabilities.

3.1 True probabilities

Second-order probabilities are often seen as probabilities that the *true* probability of something takes a particular value: "“True” may be interpreted as the value that would be assigned if certain information were available, including information from reflection, calculation, other people, or ordinary evidence" (Baron, 1987). This interpretation, though quite common and intuitive, does not itself suggest how second-order probabilities originate. In section 2.3 they describe the possibility of belonging to a particular small world. Theoretically these probabilities arise from an observation of the mass of a particular small world (group) in the whole sample.[5] Yet, in a one-shot game these observations are unavailable (even if one could identify the small worlds themselves). In this section we employ the properties of decision weights of ambiguity-neutral subjects to derive the "true" second-order probabilities, and subsequently use them to derive probability weights for subjects non-neutral to ambiguity.

Assume that there is a subgroup of subjects who behave in line with the hypothetical "objectively given" probability. In this case the weighting function measures the deviation of the whole sample from the behavior of this subgroup. A linear weighting function $\psi_i(p_i) = p_i$ in the prospect theory (in the gains domain) implies that the subject conforms with the EU-paradigm. For cumulative probabilities (cumulative

[5] This suggests a possible perspective on second-order probability formation through Bayesian updating, yet this is not in the framework discussed in the current paper.

prospect theory and RDU), this is a necessary and sufficient condition:

Proposition 1 *If* $\psi_i(0) = 0$ *and* $\psi_i(1) = 1$ *then* $\psi_i(p_i) = p_i$ *if and only if* i *is expected utility maximizer.*

Proof. If i is EU-maximizer then his/her decision functional is linear in probabilities $\sum_{\pi}(a + b \cdot \pi_s) \cdot u(c_s)$, which jointly with $\psi_i(0) = 0$ and $\psi_i(1) = 1$ implies $a = 0$ and $b = 1$, i.e. $\psi_i(\pi_s) = \pi_s$. If in (8) $\psi(\mu_j) = \mu_j$, which is the probability of the degenerate lottery with $\pi_s = 1$ for $s = j$ and $\pi_s = 0$ for $s \neq j$, hence we can write $V(c_1, ..c_S) = \sum_{s=1}^{S} \mu_s \cdot \phi(u(c_s))$. It follows that the DM is EU-maximizer. ■

The weighting of second order probabilities can be therefore seen as a deviation from the weighting function ascribed by an EU-subject (ambiguity-neutral). Our task will be therefore to identify the second-order probability distribution implied by the choices of ambiguity-neutral subjects and then find the corresponding weighting functions implied by the choices of other participants.

3.2 Quantification of qualitative data

Elicitation of the probability distributions follows Carlson and Parkin (1975) who suggested a method of quantification of qualitative data. In the original study, the method is used to derive inflation expectations from qualitative answers to the question: will prices fall or rise? It can be effectively applied to the estimation of the expected value of any random variable from qualitative answers to the question: will the variable take a value above the threshold or below it? (see Fig. 2). The method is based on two components: the two-parametric pdf to be estimated, and the size of the *difference limen*, i.e. the interval in which subjects are indifferent between the two qualitative options.

In the Ellsberg experiment, subjects' answers are qualitative: "urn A", "urn B" or "indifferent between the two". We observe the answers of individual subjects, each of them is supposed to belong to a particular group j with a probabilistic beief π_j. Indiidual subjects choose A iff $\phi(\pi_j) > \phi\left(\frac{1}{2}\right) \Leftrightarrow \pi_j > \frac{1}{2}$ (opposite signs for choosing urn B). To capture indifference, introduce the difference limen ε: if $\frac{1}{2} - \varepsilon \leq \pi_j \leq \frac{1}{2} + \varepsilon$

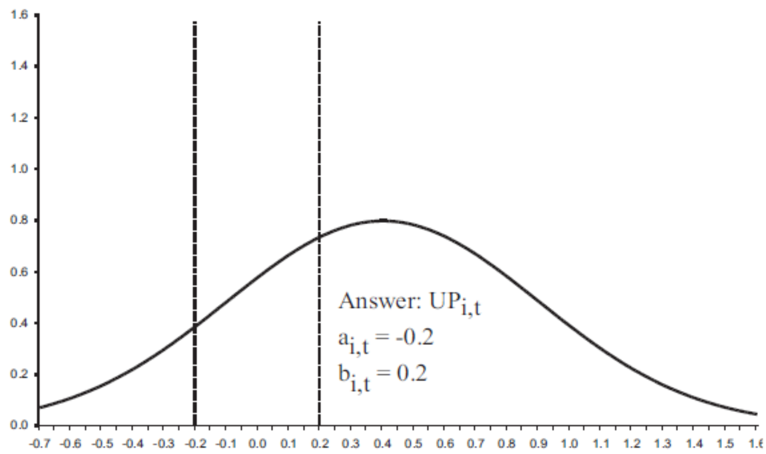

Figure 2. Quantification of qualitative data: by fixing a two-parametric pdf and the "difference limen" one determines the parameters of the distribution from the observed fractions of qualitative answers "below the threshold", "above the threshold" or "indifferent" (falls within the difference limen). Reproduced from Carlson and Parkin (1975).

then the subjects in group j cannot distinguish between the two urns. The decision of the average subject is based on the average belief as defined in (9) and depends on the weights $\psi\left(\mu_j\right) = \psi_j$ ascribed to probability distribution π_j specific to group j. Our task is to derive distribution ψ from the data. To do this note that ψ_j effectively describes the chance of the average decisionmaker to end up in group j. Fraction s_A corresponds to the probability of drawing a value above the threshold; similarly for s_B. For the average decision-maker, the total weight of the small worlds with probability measures above or below the threshold is:

$$\sum_{j:\pi_j>\frac{1}{2}+\varepsilon} \psi_j = \Pr\left(\pi > \frac{1}{2}+\varepsilon\right) = s_A,$$

$$\sum_{j:\pi_j<\frac{1}{2}-\varepsilon} \psi_j = \Pr\left(\pi < \frac{1}{2}-\varepsilon\right) = s_B.$$

In order to estimate a probability density function (pdf) of π from the above two conditions, we need to choose a two-parametric family of pdf. An additional restriction is that the pdf is defined on the interval $[0,1]$ because we are estimating second-order probabilities. An appropriate candidate is beta-distribution,[6] which meets these two

[6] The following properties of the beta-distribution can be useful. First, the $\frac{\alpha}{\beta}$−ratio fully determines the mean $\mu = \frac{\alpha}{\alpha+\beta}$. Second, for a fixed value of $\frac{\alpha}{\beta}$ (preserving the mean), the variance $\sigma = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

conditions and additionally provides the flexibility of virtually all shapes, hence the choice of the function is not very restrictive. [7] We therefore assume $\pi \sim Beta\left(\alpha,\beta\right)$. With experimentally observed s_A and s_B, we solve the following system of equations for α and β:

$$s_A = \frac{1}{\mathrm{B}\left(\alpha,\beta\right)}\int\limits_{\frac{1}{2}+\varepsilon}^{1}\pi^{\alpha-1}\left(1-\pi\right)^{\beta-1}d\pi, \tag{10}$$

$$s_B = \frac{1}{\mathrm{B}\left(\alpha,\beta\right)}\int\limits_{0}^{\frac{1}{2}-\varepsilon}\pi^{\alpha-1}\left(1-\pi\right)^{\beta-1}d\pi. \tag{11}$$

The above system of equations has a unique solution (α,β). This is because for any value of s_A and s_B the equations define implicit functions $\beta=\beta_{S_A}\left(\alpha\right)$ and $\beta=\beta_{S_B}\left(\alpha\right)$ with strictly positive but different slopes and there exists a unique intersection point. The solution is found numerically on the grid $\alpha,\beta\in\mathbb{N}$ as $\alpha = \arg\min\left(\beta_{S_A}\left(\alpha\right)-\beta_{S_B}\left(\alpha\right)\right)^2$. This choice of the grid is sufficient in most cases as typical values of α and β are above 10.[8]

Once we have identified α and β for the sample, we can derive the weights $\psi_j=\Pr\left(\pi\leq\frac{j}{100}\right)-\Pr\left(\pi\leq\frac{j-1}{100}\right)$, the mean $\overline{\pi}=\sum\limits_{j=1}^{100}\psi_j\cdot\pi_j=\frac{\alpha}{\alpha+\beta}$ and the variance $\sigma_\pi^2=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.[9] The latter represents the variance across the whole sample. [10] The mean is the sample's average belief with regards to π, however it should not be

---

decreases both in α and β. It turns out that an inverse-S shaped probability weighting function is more likely to arise if the implied pdf for EU-subjects is rather flat (low values of α and β) and the pdf for non-EU subjects is "tall and thin" (high values of α and β), otherwise (comparable values of parameters in both pdf) small variations in parameters lead to significant changes in the resulting shape of the weighting function.

[7] Other pdf families could include, for example, trancated normal, logit-normal, Kumaraswamy distribution, among others. As we are not looking for a best-fit distribution but rather aim to derive the first moment from qualitative data, we choose the one that offers the easiest implementation. It is worth noting that generalized weighting functions used in the prospect theory are also suitable two-parametric pdf families: $\frac{\alpha\pi^\beta}{\alpha\pi^\beta+(1-\pi)^\beta}$ as in Lattimore et al. (1992) and $e^{-\beta(-\ln\pi)^\alpha}$ in Prelec (1998). These functions are relatively well studied as weighting functions but not as pdf. An additional advantage of using beta-distribution is that it is *not* usually linked to the probability weighting functioins, and hence is not expected to generate the saught result.

[8] For lower values we improve precision to the first decimal number.

[9] This choice of the pdf family is advantageous also because it allows one to easily model benchmark distribution with given parameters $\overline{\pi}$ and σ_π by assigning $\beta=\alpha\left(1-\overline{\pi}\right)/\overline{\pi}$ and $\alpha=\left(1-\overline{\pi}\right)\cdot\overline{\pi}^2/\sigma_\pi^2-\overline{\pi}$. This becomes relevant if one wishes to model the hypothetical "true" distribution carried by the signals.

[10] To further clarify the notion of the "average decision-maker", recall that all subjects act as if each (group) of them, denoted with j, has a specific primary probability π_j in mind. Then $Beta\left(\alpha,\beta\right)$ is the distribution of fractions s_j of these (groups of) subjects. The "average decision-maker" can pick any of these primary distributions with probability s_j. Hence for him s_j is effectively the second-

regarded as the belief of the "average decision-maker" because the latter is defined as second-order utility maximizer and has to base his decision on (9) which is the average value of $\phi(\pi_j)$; this average is then compared with $\phi\left(\frac{1}{2}\right)$ to yield the choice between urn A and B.

4 Experiment

The experiment is run as an online survey mimicking the original two-urn Ellsberg task and its variations (see below). Typical results from an Ellsberg experiment reveal about 10-30% of ambiguity-neutral subjects (e.g. Akay et al.,2010, report 57% AA, 24% AN and 20% AL). Since ambiguity-neutral subjects (EU-maximizers) play an important role in this study (their choices are used to derive the benchmark "true" distribution μ_j), we need to obtain a large enough subsample of them. Apart from the usual choice options (urn A or B), we also need to include an "indifference" option to reflect the difference limen assumption: upto 40% subjects choose "indifferent" as an answer to a choice question.[11] The required sample size is therefore hard to achieve in a lab experiment (with typical group sizes of 20-40 participants), therefore most of the data is collected from an online survey, see Table **??**. For comparison, in a similar experiment with 274 participants (MBA students) HE(1986) report 34% subjects inddiferent between the two options, and 77% of the rest choosing the risky urn (urn

order probability of primary probability π_j. This gives an intution of the notion of the average subject and why $Beta(\alpha, \beta)$ is the distribution of his second order probabilities. The variance of $Beta(\alpha, \beta)$ characterizes the dispersion of the primary probabilities. If subjects j act as SmEU-maximizers, then π_j should be seen as the mean values of their respective distributions, fractions s_j as probabilities of being in group j, i.e. probability of having the respective probability distribution in mind, and the variance of $Beta(\alpha, \beta)$ as the variance of these means. The variance of $Beta(\alpha, \beta)$ aggregates underlying variances but is not necessarily equal to the average variance of the underlying second-order distributions of sugjects j. In fact, it is typically larger than individual variances as the "average decision-maker" can with certain probabilities belong to the group with the smallest mean, and with the largest mean, as well to all groups inbetween, which potentially increases the support of his distribution. The variance can only be equal to the average variance, if the distribution of the "mean" group "covers" all other distributions.

[11] Indifference, generally, is hard to interpret in experiments. In the Ellsberg task indifferent subjects could be treated both as EU-subjects with underlying probabilities 50-50 in mind, or as non-neutral to ambiguity subjects that do not perceive the difference between the two choices to be significant (for whatever reason) and therefore randomly choosing the answer. It is also possible that subjects choose the indifference option if they are reluctant to give answers. In our survey, subjects have an exit option and therefore we expect the majority of "indifferent" answers to fall in one of the above categories. In order to identify the ambiguity attitude of indifferent subjects, we additionally ask them to answer the same question without the indifference option, see below.

B in our notation); in TVW(2008) 73% participants (out of 63) prefer the risky urn, and in KP(2003) 75% (out of 61) prefer the risky urn. Conducting online surveys has allowed us to reach the numbers of participants in the range 500-1000.

We consider treatments with and without monetary incentives, as well as conduct a robustness check through equivalent in-class treatments. An important feature of our experimental design is the exit option at any stage of the survey (participants can close the on-screen window and quit the system but their answers to that point would be recorded; the completion rate in online treatments is about 70%). This allows us to control for non-monetary motivation: subjects who completed the whole survey on average produce results closer to those obtained in lab or iin-class sessions. Shadrina and Vinogradov (2013) show that in these treatments non-monetary incentives are strong enough to outperform monetary factors: the latter become insignificant once non-monetary factors are controlled, yet remain significant without controls for non-monetary motivation. Without entering the discussion of whether lab results with usual controls are more or less relevant to the real life behaviour than the results from an online survey[12] (see e.g. Rubinstein, 2012, for some discussion), we present data from several independent online treatments that are consistent with each other and with other experiments and hence deserve attention.

The whole experiment consists of four parts: A, B, C, and D. In the beginning (Part A), the standard Ellsberg task is described. The description of the original set up is available to participants throughout the experiment. The full questionnaire with exact formulation of the questions is in the Appendix. Part A is used to identify ambiguity averse, ambiguity liking and ambiguity neutral (EU) subjects: a subject is called ambiguity averse if (s)he chooses urn B when asked to draw a black ball, and the same urn when asked to draw a red ball (similarly for ambiguity liking and neutral). It also serves as a basis for comparison with other Ellsberg experiments.

In addition to each choice question in Part A ("which urn would you choose if you were offered a prize if red (black) ball is drawn from your chosen urn"), we also ask

[12] Advantages of online experiments would include (a) no pressure on participants, (b) natural environment, (c) no biases through being observed by others (e.g. fear of negative evaluation), (d) flexibility in time etc.

	2011 online			2012 class			2012 online		
	N	A(Red)	B(Red)	N	A(Red)	B(Red)	N	A(Red)	B(Red)
	765			49			487		
A(Black)		.200	.050		.211	.053		.255	.029
B(Black)		.059	.692		.175	.561		.101	.616

Table 1. Ambiguity attitudes. Only fully completed questionnairs.

a paired question "what would be your choice if you had no indifference option". This allows us to identify which fraction of indifferent subjects falls into ambiguity averse, ambiguity-loving or EU category when *forced* to make a decision. In a total sample of over 700 participants we had 199 subjects indifferent between urn A and B when answering the first type of questions. When answering the second type of questions, only 23 of them give answers consistent with the EU-paradigm (either choosing urn A when asked to bet on a red ball and urn B when asked to bet on the black, or vice versa), whereas 162 exhibited non-neutrality to ambiguity (102 ambiguity averse, and 60 ambiguity-loving).[13] Table 1 summarizes the ambiguity attitudes of the subjects in our surveys as revealed by questions in Part A.

Part B consists of questions testing subjects' responses to hypothetical signals. The first two signals communicate a hypothetical choice of other participants (signal S1 gives the absolute number of other participants with this choice, and signal S2 - the fraction of them). Both signals are irrelevant for the decision-making as they do not communicate directly anything about the distribution of balls in the ambiguous urn. The three other signals communicate the proportion of other players who have picked a red ball from the ambiguous urn (in S3 and S5 this fraction is 60% but the communicated sample size in S5 is larger; in S4 the communicated fraction is 80%). These signals are relevant for decision-making as they provide information on the possible distribution of balls in the ambiguous urn, although do not remove ambiguity. The numbers in the signals are chosen so that it is easy to calculate the signalled ratios and compare them with each other.

Part C measures subjects' confidence in their choices by asking them whether they would draw from a different urn if they pick the ball of a wrong color and whether

[13] These fractions result in a rather "flat" pdf for EU-subjects (benchmark distribution) and "tall and thin" pdf for the whole sample, as well as for ambiguity averse and ambiguity loving subjects separately.

	Signals (questions Q3-Q7)					
	Initial	S1	S2	S3	S4	S5
s_B	.495	.556	.536	.433	.374	.372
s_A	.130	.212	.253	.438	.523	.503
s_{indiff}	.395	.252	.210	.128	.103	.124
N	397	397	397	397	397	397
Implied distribution of $\pi \sim Beta\,(\alpha, \beta)$ for $\varepsilon = .02$:						
α	99	34	28	8	8.2	10.2
β	107	38	31	8	7.4	9.4
Mean	.480	.472	.475	.500	.526	.520
St.dev	.0347	.0584	.0645	.1213	.1226	.1100

Table 2. Fractions of answers A, B, or indifferent to questions Q1-Q5 in 2012 online survey.

they would do so if they know in addition that the majority of other subjects have switched to a different urn. Part D is reserved for demgraphic information and self-assessment of the subjects' proficiency in statistics. The role of parts C and D in the current paper is to control for non-monetary incentives by identifying the subsample of fully completed surveys.

5 Results

The observed fractions s_A, s_B and s_{indif} of subjects choosing urn A, B or the indifference option respectively as their answers to questions from part A and part B are used to derive the distributions of weighting coefficients ψ_j as described in section 3.2. To do this, we filter the data from the three surveys with the indifference option by removing incomplete questionnaires and those with inconsistent answers.[14]

5.1 Distributions

To obtain a solution for (10-11) we set the difference limen at $\varepsilon = .02$. Table 2 provides a summary of results for the 2012 online survey and Fig 3 depicts the resulting second-order probability distributions.

The choice of the difference limen at $\varepsilon = .02$ is similar to that in Carlson and Parkin (1975) who assume that subjects are insensitive to changes in prices under 2%.

[14] An answer is deemed inconsistent, if a subject chooses urn A to draw a red ball in part A but subsequently switches to urn B in one of the questions in part B: signals in part B are constructed to communicate that the chance of drawing red from A is *at least not worse* after the signal than before.

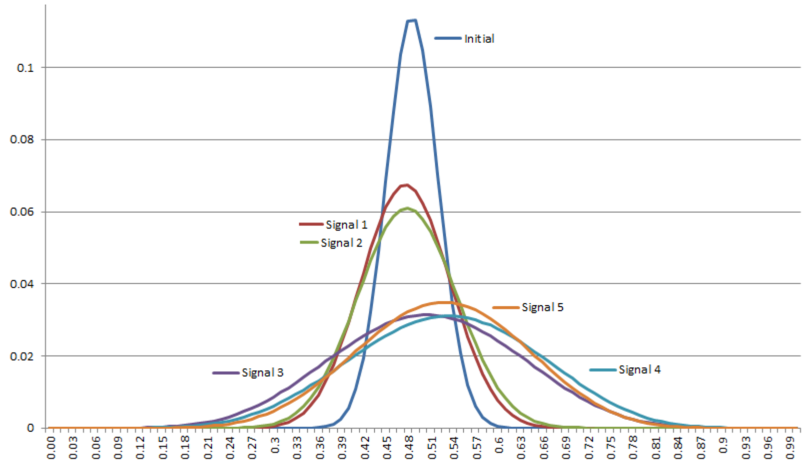

Figure 3. Effects of signals on probability distributions of π. Sample: $N = 397$ valid observations, online, 2012. Difference limen $\varepsilon = .02$.

There are no comparable studies on sensitivity to differences in probabilities. Kahneman and Tversky (1979) note that subjects can be insensitive to a difference between "very high" and "very low" probabilities and unity or zero respectively. We believe that 2% is a reasonable assumption. The choice of the difference limen effectively changes the "thickness" and the "height" of the distribution (see Fig. 4 for comparison), yet affects all distributions simultaneously and equally. In this paper we are interested in the relative effects, which are unaffected by the choice of ε.[15]

In part B signals S1 and S2 (on the *preferences* of other subjects) were initially included to ensure that subjects are well prepared to understand the "main" signals S3-S5 (on the *observations* from draws from urn A by other subjects). We regard the information in S1 and S2 as irrelevant for the decision-making, and therefore expected no effect of these signals on the subjects' choices. Yet, table 2 and Fig 3 reveal a noticeable impact of these signals, though quite different from that of S3-S5.

It is worth noting that all signals increase the standard deviation of the second-order distributions, except for S6 (60% chance of drawing red from A, as in S4, but based on 200 observations instead of 20, i.e. S6 has a higher precision). This is because

[15] We tested several values for the difference limen. If the values are too low or too high, the solution of equations (10-11) produces either I-shaped (very thin and high) or U-shaped distributions because the difference limen has to accommodate the fraction of answers "indifferent", which varies in size for different signals. Setting $\varepsilon = .02$ also tackles this problem and allows us to consistently use the same difference limen in all subsamples and for all signals.

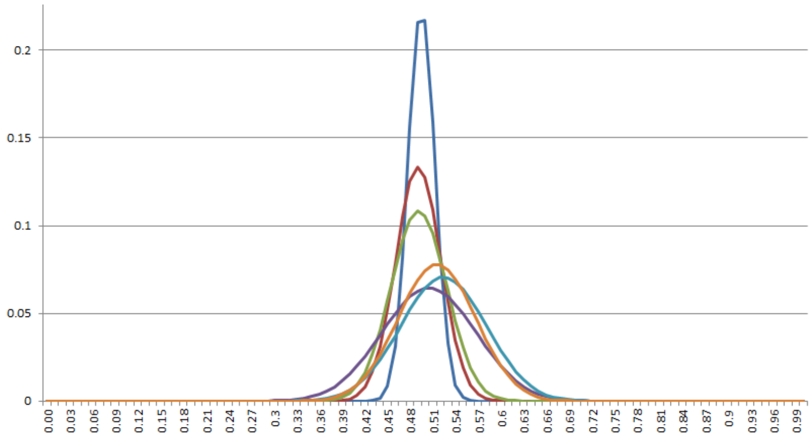

Figure 4. Same as Fig. 3 but with difference limen $\varepsilon = .01$.

the increase in the fraction s_A of those who choose urn A happens to a larger degree at the expense of the fraction of indifferent subjects, and to a smaller degree at the expense of those who prefer urn B. This indicates the presence of a non-negligeable fraction of ambiguity-averse subjects who require a much stronger signal to switch from urn A to urn B, close to maxmin decision-makers of Gilboa and Schmeidler (1989). [16]

5.2 (Second-order) Probability weighting

Subjects are identified as ambiguity-averse, ambiguity neutral (EU) and ambiguity lovers, as described in section 4.[17] For each of these three subsamples we obtain distributions of π, similar to those Fig. 3 for the whole sample. The resulting parameters are in Table 3.

To obtain the probability weighting, we use the fact EU-subjects assign weights $\psi(\mu_j) = \mu_j$ to probability measures π_j. These values are now known from the pdf function with parameters α and β obtained for the subsample of EU subjects. A similar pdf for non-neutral to ambiguity subjects produces values ψ_j for their weights. By combining the two, we obtain the probability weighting of non-neutral subjects as $\psi(\mu_j) = \psi_j$, which is plotted in Fig. 7. By using the data from Table 2, a similar

[16] From the perspective of small worlds, used in this paper to interpret second-order probabilities, after the signals we observe an increase in the mass of small groups at the lower end of the distribution. This is somewhat puzzling and deserves a further investigation.

[17] Ambiguity attitude is identified from questions without an indifference option (Q1* and Q2*) and assumed to remain unaffected by signals. Therefore we know the responses of subjects with different ambiguity attitudes to all our questions.

		Signals (questions Q3-Q7)					
		Initial	S1	S2	S3	S4	S5
EU	s_B	.500	.395	.526	.342	.447	.526
	s_A	.161	.368	.316	.526	.474	.395
	α	149.2	28.4	13	9.8	3.4	3.2
	β	161.6	28.6	14.4	8.8	3.4	3.6
	Mean	.480	.498	.474	.527	.500	.471
	St.dev	.028	.066	.094	.113	.179	.179
AA+AL	s_B	.481	.565	.553	.435	.365	.363
	s_A	.161	.204	.238	.442	.536	.514
	α	80	31.6	22.2	7.8	7	9
	β	86	35.6	25	7.8	6.2	8.2
	Mean	.482	.470	.470	.500	0.530	0.523
	St.dev	.039	.060	.072	.123	.132	.117

Table 3. Parameters of distributions for ambiguity neutral and non-neutral subjects (difference limen .02). Survey: online, 2012.

excercise produces a probability weighting function for the whole sample, see Fig.8.

We have two options. First, we can estimate non-cumulative weits $\psi\left(\mu_j\right)$ as in the original prospect theory. Yet, in addition to the traditional stochastic dominance criticism of this approach we face the following difficulties: (1) the values of probabilities observed in figures 3 and 5,hardly exceed 0.25, which results in a potentially too narrow range of weight estimates, and (2) as examplified in Fig. 6, in the same range of second-order probabilities (y-axis), we can observe both overweighting (weights assigned by non-neutral to ambiguity subjects are above those of ambiguity neutral ones) and underweighting, depending on whether we consider the lower or the upper parts of the primary probabilities scale (x-axis).[18]

The second option is to apply the probability weighting function to the *cumulative* probabilities by using $\psi\left(\mu_j\right) = \psi^C\left(\sum_{k=1}^{j}\mu_k\right) - \psi^C\left(\sum_{k=1}^{j-1}\mu_k\right)$. For the subsample of ambiguity neutral subjects we use the elicited values of $\alpha = \alpha^{EU}$ and $\beta = \beta^{EU}$ to calculate $\sum_{k=1}^{j}\mu_k = \Pr\left(\pi \leq \pi_j \middle| \alpha^{EU}, \beta^{EU}\right)$, and for the subsample of non-neutral subjects use $\alpha = \alpha^{NN}$ and $\beta = \beta^{NN}$ to calculate $\sum_{k=1}^{j}\psi_k = \Pr\left(\pi \leq \pi_j \middle| \alpha^{NN}, \beta^{NN}\right)$ to

[18] This effect is less visible in Fig. 5 although also present in the range of ψ of 0.1-0.15.

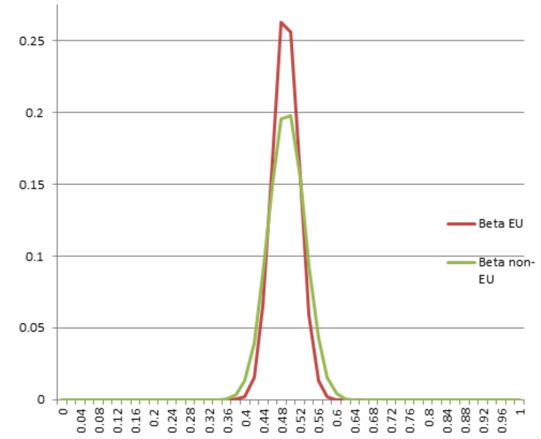

Figure 5. Second-order probability distributions for EU and non-EU subjects resulting from the original Ellsberg task.

obtain the mapping $\sum_{k=1}^{j} \psi_k = \psi^C \left(\sum_{k=1}^{j} \mu_k \right)$:

$$\frac{1}{\mathrm{B}\left(\alpha^{NN}, \beta^{NN}\right)} \int_0^{\pi_j} \pi^{\alpha^{NN}-1} (1-\pi)^{\beta^{NN}-1} d\pi = \psi^C \left(\frac{1}{\mathrm{B}\left(\alpha^{EU}, \beta^{EU}\right)} \int_0^{\pi_j} \pi^{\alpha^{EU}-1} (1-\pi)^{\beta^{EU}-1} d\pi \right).$$

The resulting [cumulative] probability weighting function ψ^C has a usual inverse-S shape, typical for probability weighting functions in the prospect theory: it demonstrates an overweighting of low probabilities and underweighting of high probabilities.

As in the prospect theory (see Appendix A), the weighting function reflects ambiguity attitude of the subjects. Yet, in the original prospect theory ambiguity attitude is captured by the weighting of non-cumulative probabilities, whereas there exists no mapping of cumulative probabilities to weights that would reflect ambiguity attitude of the subjects.[19] In the second-order utility model the weighting of cumulative probabilities is similar to that in the cumulative prospect theory, as long as the majority of subjects exhibit amibuity aversion.[20]

Ambiguity aversion is typically associated with pessimism, and ambiguity-loving

[19] As everywhere throughout the paper, we assume that the system of weights adds up to unity. It is possible to have a subadditive system of weighing coefficients that reflects ambiguity aversion, as done in Choquet expected utility with the help of subadditive capacities.

[20] In our sample, as usual, the majority of subjects are ambiguity averse, therefore we cannot construct an estimate of the pdf and related weighting function for a sample with a majority of ambiguity lovers. It does not make much sense to split the subsample of non-EU subjects in AA and AL subsamples, as for these subsamples either $s_A = 0$, or $s_B = 0$. However, we construct a hypothetical sample with a majority of ambiguity lovers in Fig. 10.

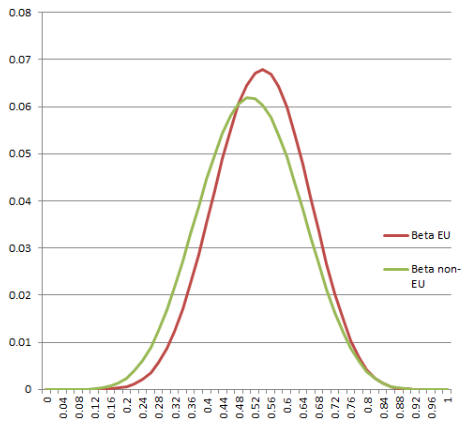

Figure 6. Second-order probability distributions for EU and non-EU subjects after signal S3 (60% chance to get Red from urn A, based on 20 observations).

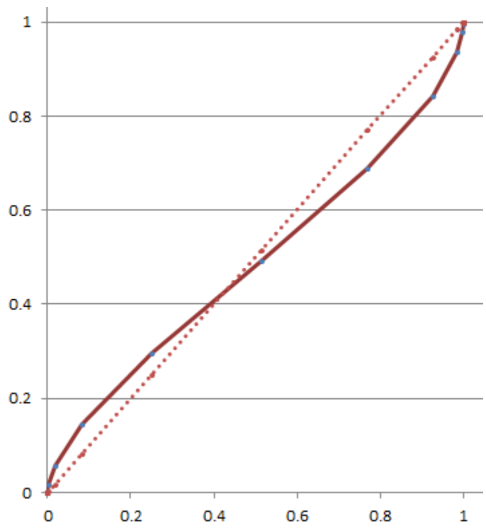

Figure 7. Second-order probability weighting function for ambiguity-non-neutral subjects. Survey: online, 2012.

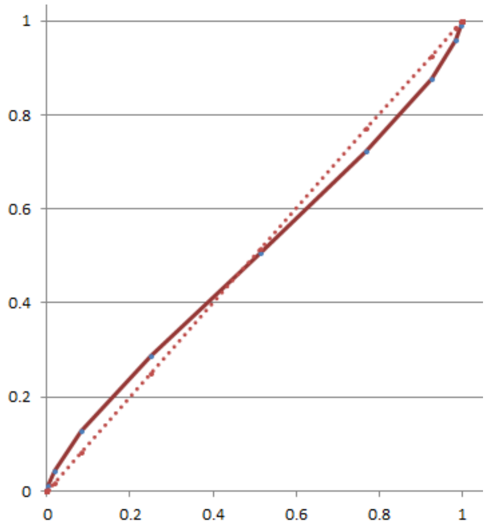

Figure 8. Second-order probability weighting function for ambiguity-non-neutral subjects. Survey: online, 2012.

behavour - with optimism. Viewing weighting functions from the optimism-pessimism perspective, as in Fig. 1 (Weber, 1994) is based on a rather straightforward intuition: a pessimistic (optimistic) subject assigns higher weights to lower (higher) ranked prospects, resulting in a concave (convex) weighting function. This approach explicitly assumes that subjects disregard information communicated by the objective probability distribution, and instead behave as if a different (more pessimistic, or optimistic) probability distribution was communicated. This can be interpreted as if subjects allow for a degree of imprecision associated with the communicated distribution, in which case optimism and pessimism obtain the ambiguity attitude interpretation usual for the literature on Knightian uncertainty. From the perspective of SOEU, optimistic subjects assign higher weights to the first-order probability distributions that offer a higher expected payoff (see Fig. 9), which results in a visibly more pronounced convex part of the weighting function (Fig. 10). The standard inverse-S shape is therefore due to the majority of ambiguity averse subjects in the sample.

As a next step, we ask how signals affect probability weightings used. In fact, if ambiguity attitude is unaffected by signals (which is plausible), the only way signals can change decisions, is by changing the (second-order) weights ascribed to the primary probabilities (small worlds).

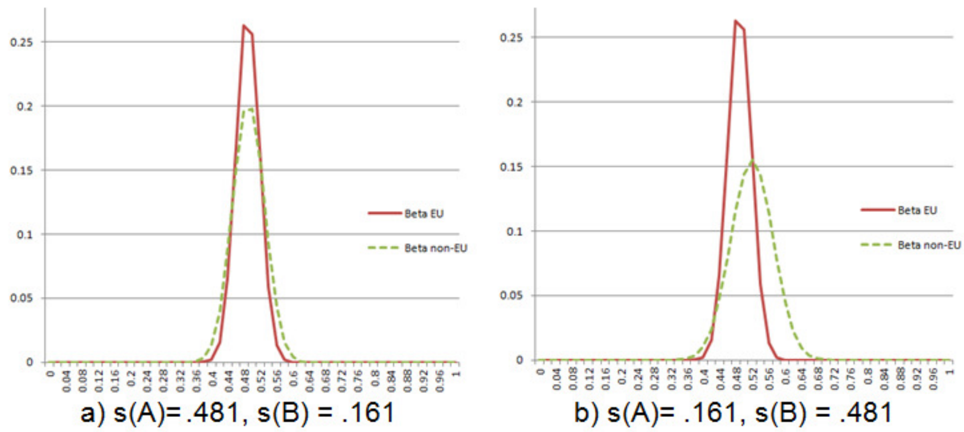

Figure 9. Second-order probability distributions for ambiguity-neutral (EU) and non-neutral (non-EU) subjects: a) experimental data with a majority of ambiguity averse subjects, (b) reverted fractions s_A and s_B to a hypothetical example of a sample with a majority of ambiguity-lovers.

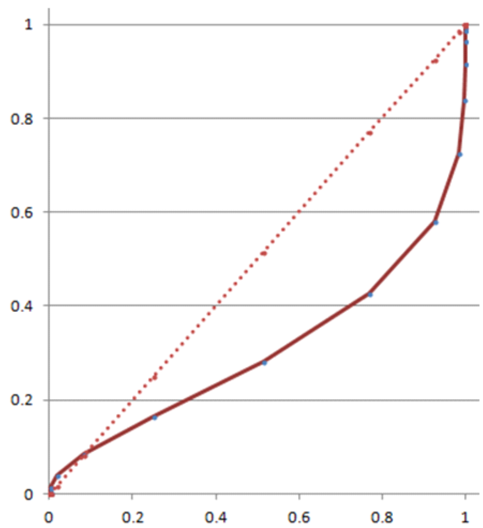

Figure 10. Weighting function for a hypothetical sample with a majority of ambiguity lovers.

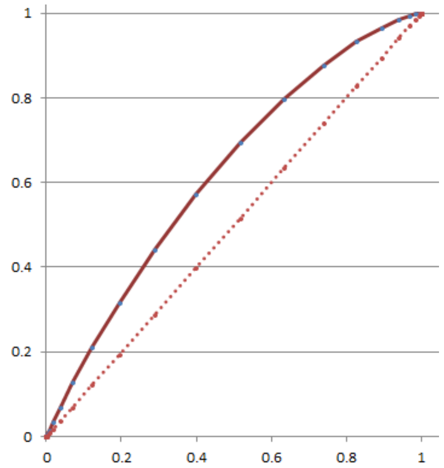

Figure 11. Second-order probability weighting function for ambiguity-non-neutral subjects - after signal S1 (non-probabilistic).

5.3 Effect of signals: lost inverse-S shape

A potential concern with the approach used in this paper is that the choice of the particular probability distribution function might generate the result in the previous section. However, this is not the case: in fact, the inverse-S shape of the weighting function is vulnerable to distortions. In our survey, such distortions are generated by signals that are initially designed to affect subjects' ideas of a possible distribution of balls in urn A. Figure 11 demonstrates that a rather irrelevant signal ("12 subjects before you have chosen urn A to draw a Red ball") makes the weighting function look like if non-neutral to ambiguity subjects overestimate probabilities on the whole range of them (pessimistic shape). A similar picture is obtained in Fig. 12, as a reaction to the signal "12 out of 20 subjects have drawn Red from A."

The reason for this is that EU-subjects (ambiguity neutral) who serve as the benchmark for the derivation of the probability weighting function, react to the signals stronger than ambiguity-neutral ones, as shown in Fig. 6.[21] Recall that we only consider average decision-makers; the average decision-maker for the ambiguity-neutral subsample aggregates beliefs of all small groups formed by ambiguity-neutral subjects.[22] In the framework of small worlds (small groups) employed in this paper,

[21] The differences in the reaction fo ambiguity-neutral and non-neutral subjects to news (signals) are analyzed in details in Vinogradov (2012).

[22] Each small group operates in a small world that implies a unique probability measure. However

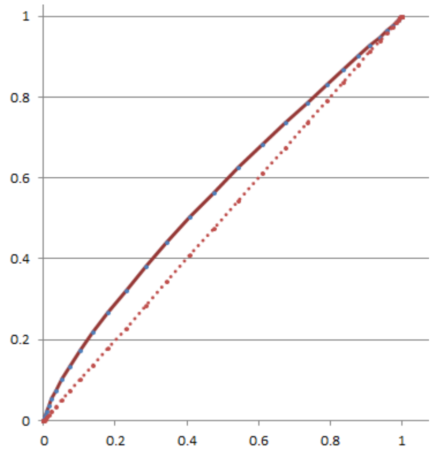

Figure 12. Second-order probability weighting function for ambiguity-non-neutral subjects - after signal S3 (probabilistic).

such an effect of a signal corresponds to a major revision of the probability measure by ambiguity-neutral subjects, and hence them moving from one small group to another, and a minor or no revision of the probability measure by those non-neutral to ambiguity.

The reaction of EU-subjects to the signals S4 and S5 is also remarkablly different from that of non-EU ones, see the second-order distributions and the resulting weighting functions in figures 13-16.

ambiguity neutral and non-neutral subjetcs take different factors into account and hence are in different worlds, even though occasionally these worlds can produce equal probability measures. Hence, a signal that changes the world (and the probability measure) for a group of ambiguity-neutral subjects, does not need to change the world for those non-neutral to ambiguity, and vice versa.

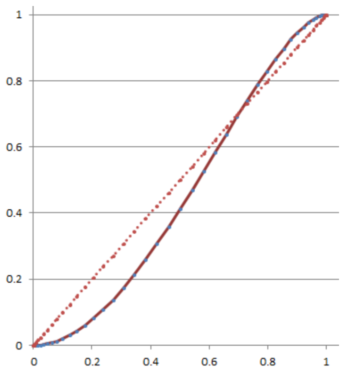

Figure 13. Second-order probability weighting function for AA and AL subjects after signal S4 (16 out of 20 positive observations).

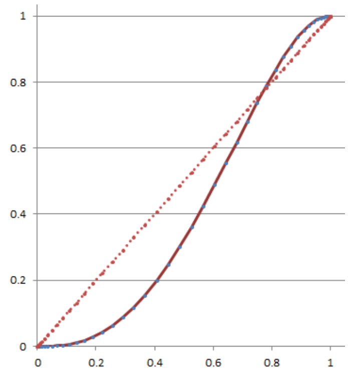

Figure 14. Second-order probability weighting function for AA and AL subjects after signal S5 (120 out of 200 positive observations).

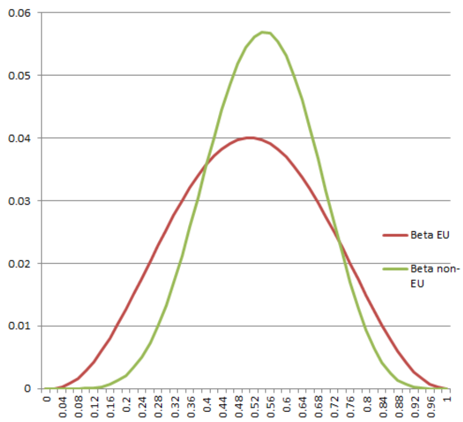

Figure 15. Second-order probability distributions for EU and non-EU subjects after signal S4 (80% chance to get Red from urn A, based on 20 observations).

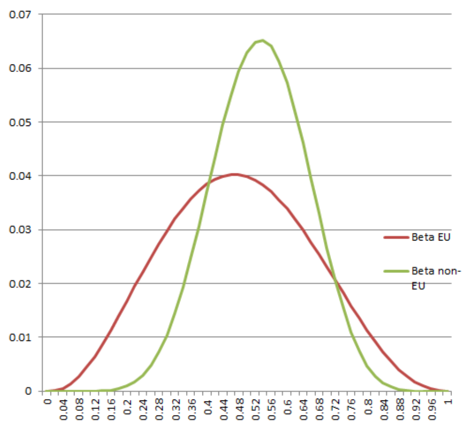

Figure 16. Second-order probability distributions for EU and non-EU subjects after signal S5 (60% chance to get Red from urn A, based on 200 observations).

5.4 Modelling the "true" distribution: inverse-S restored

In the above analysis the implied second-order probabilities of EU-subjects were taken as a benchmark to derive the weighting functions. In a companion paper, where the focus is primarily on the effects of signals on the decisions in the Ellsberg experiment, we show that it is the EU-subjects who tend to react to irrelevant news (this result is significant at 5% level). However, once they are communicated a probabilistic signal, they fail to act exactly as prescribed by the communicated probability: although we would expect them to have the mean of the distribution around 60% and 80% level when receiving signals S3 (S5) and S4 correspondingly, and vary the standard deviation of the distribution in response to the precision of the signal as given by S3 and S5, figures 6, 15 and 16 reveal that the mean of the second-order distribution does not move as expected. Instead, it remains close to the initial 50%. This is in line with the observations of Mankiw et al. (2003) who demonstrate that the distribution of inflation expectations does not fully incorporate the policy change signal; instead, the density kernel of expectations moves gradually: people exhibit inertia in processing news.

This brings us back to the discussion of "true" probabilities from Section 3.1. So far it was assumed that the true probability distribution is given by the one obtained for EU-subjects. The reason for this was the assumption that EU-subjects assign weights $\psi\left(\mu_j\right) = \mu_j$ to second-order probabilities μ_j. Yet, if this assumption is relaxed, we have to find a different way of eliciting the true distribution.

As a first attempt, we take the initial probability distribution of the EU subjects as a benchmark, and map probabilitis implied by signals as weighting functions, see figure 17. This exercise confirms that in response to the most precise (as given by the hypothehsized number of prior observations 200) probabilistic signal, EU-subjects demonstrate an even higher degree of pessimism[23] than AA-subjects. This contradicts to the intuitive meaning of a signal that communicates that approximately 60% of balls in the ambiguous urn should be of the red colour, and hence represents an optimistic

[23] This term is used as in Weber (1994), in its usual application to rank-dependent outcomes: subjects assign higher weights to lower ranked outcomes and vice versa.

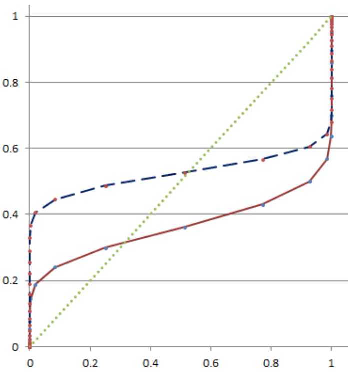

Figure 17. Implied ψ for EU (dashed line) and AA+AL (solid line) subjects after signal S5 (120/200 positive observations).

signal as compared to the initial belief of 50% red balls.

It follows that the "distorted" probability distributions of EU-subjects are not a suitable proxy for the "true" distribution. Instead, we model the true distribution by making the following assumptions. First, the true distribution should have the mean that corresponds to the mean communicated by the signal. For signals S3 and S5 this will be 60%, and for signal S4 it is 80%. Second, the variance of the true distribution should be consistent with the variance of the initial beliefs of the EU subjetcs, as in Table 3. Third, more precise signals should correspond to true distributions with smaller variances. Fourth, the true distribution belongs to the same family of PDF as elicited distributions (in our case this is beta-distribution). With this in mind, we construct the weighting functions for second-order probability distributions obtained after signal S5 in figure 18. In this excercise, the weighting function of non-neutral to ambiguity subjects restores to the one obtained for the initial measurement in figure 7. As above, the figure reveals a higher degree of pessimism of EU-subjects, which we explain with their inertia in updating beliefs.

6 Conclusion

Explaining subjects' behaviour with second order probability distributions is useful in interpreting results from Ellsberg experiments with signals about the distribution of balls in the ambiguous urn. Ambiguity averse subjects are more confident in their decisions than EU-maximizers; the latter tend to "desperately seek" for a "true"

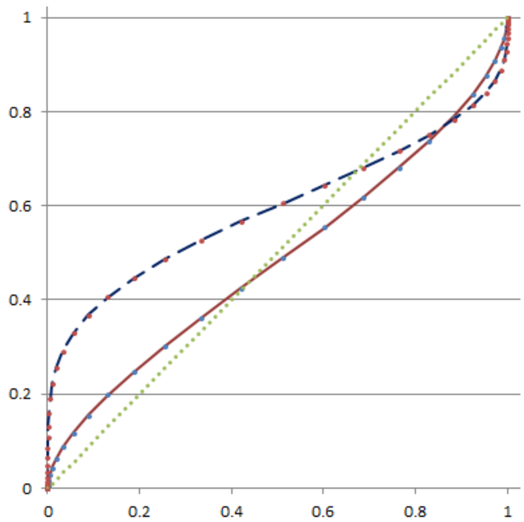

Figure 18. Weighting functions fitted to a hypothetical "true" distribution ($\alpha = 17.2, \beta = 16$) providing a close match of the weighting function for non-EU subjects to the initial weighting function as in Fig. 7.

probability distribution and react to irrelevant news. The "weighted" version of SmEU finds empirical support. In particular, it provides an intuition as to why and how people react to news and information. This approach also offers a link between the theory of decisions in ambiguity and prospect theory of decisions in risky environments: if probability weighting functions only capture subjects' inability to correctly process probabilities, this should hold both for primary (first order) and secondary (second order) probability distributions. Indeed, in the absence of signals this weighting function exhibits a standard inverse-S shape known from the prospect theory.

Signals distort behaviour of EU-maximizers stronger than that of AA-subjects, and the resulting implied weighting function does not resemble the standard inverse-S shape unless one assumes that signals do not directly enter the decision functional but rather distort the weighting of initial (undistorted) second-order probabilities. This appears to be particularly true for EU-subjects, and is consistend with the observations of inertia in updating the probability distributions of inflation expectations in response to policy changes.

Contrast to the typical assumption that EU-subjects use probabilities as decision weights, we provide evidence that they rather distort probabilities when face signals that do not match their initial beliefs. This can be seen as pessimism, yet of a different sort than that of ambiguity-averse subjects. Whilst the latter can be seen as assigning higher weights to the probability distributions that offer lower expected outcomes, EU-subjects assign lower weights to probability distributions that differ from their initial beliefs. In a sense, this can be seen as "news aversion" and can only be observed in a dynamic setting, unlike ambiguity aversion that is usually detectable in a static framework.

References

Baron J. (1987) "Second-order probabilities and belief functions," *Theory and Decision*, 23(1), pp 25-36

Carlson J., Parkin, M. (1975) "Inflation Expectations", *Economica,* New Series, Vol. 42, No. 166, pp. 123-138

Chew, S. H., and Sagi, J. S. (2008) "Small worlds: Modeling attitudes toward sources of uncertainty,".*Journal of Economic Theory*, 139(1), 1–24.

Ergin, H., and Gul, F. (2009) "A subjective theory of compound lotteries," *Journal of Economic Theory*, 144(3), 899–929.

Gilboa, I. and D. Schmeidler (1989) "Maxmin Expected Utility with Non-Unique Prior," *Journal of Mathematical Economics*, 18, 141-153.

Kahneman, D., Tversky A. (1979) "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, pp. 263-291

Kahneman D., Tversky, A. (1992) "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, 5:297–323, 1992

Klibanoff, P., Marinacci, M., Mukerji, S. (2005) "A Smooth Model of Decision Making under Ambiguity", *Econometrica,* 73(6), pp. 1849–1892

Lattimore, P., Baker, J., Witte, A. (1992) "The influence of probability on risky choice : A parametric examination", *Journal of Economic Behavior & Organization,* 17(3), pp. 377–400

Mankiw, N.G., R. Reis, J. Wolfers (2003) “Disagreement about Inflation Expectations”, NBER Macroeconomics Annual 2003, Volume 18

Nau, R. F. (2006). Uncertainty aversion with second-order utilities and probabilities. *Management Science*, 52(1), 136–145

Neilson W.S (2012) "A simplified axiomatic approach to ambiguity aversion," *Journal of Risk and Uncertainty* 41(2), pp. 113-124.

Prelec, D. (1998), "The probability weighting function," *Econometrica*, 66, pp. 497–527.

Quiggin, J. (1982), 'A theory of anticipated utility', *Journal of Economic Behavior and Organization* 3(4), 323–43

Rubinstein, A. (2012) "Response Time and Decision Making: A Free Experimental Study", mimeo, Tel Aviv University

Savage, J.L.(1954) The Foundations of Statistics, Wiley, New York.

Shadrina E., D. Vinogradov (2013) "Non-monetary incentives in online experiments", *Economics Letters*, 119(3), pp. 306–310.

Vinogradov D. (2012) "Good news: it's bad news!", mimeo, University of Essex. Presented at "Models of Expectation Formation and the Role of the News Media for Information Transmission," University of Hamburg

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

Weber, E.U., (1994) "From Subjective Probabilities to Decision Weights: The Effect of Asymmetric Loss Functions on the Evaluation of Uncertain Outcomes and Events," *Psychological Bulletin*, 115(2), No. 2, pp. 228-242

Appendix A. Prospect theory in ambiguity

Consider a decision-maker with prospect theory preferences. Facing a prospect that delivers outcomes c_s in states $s = 1..S$ occuring with probabilities μ_s, this decision-maker assigns weights $\psi(\mu_s)$ to the values $v(c_s)$ of state dependent outcomes and thus maximizes

$$V(c_1, ..c_S) = \sum_{s=1}^{S} \psi(\mu_s) \cdot v(c_s).$$

To confront the decision-maker with ambiguity, consider the choice between a risky and an ambiguous prospect, as in the two-urn Ellsberg (1961) experiment. In the original thought experiment, the decision maker is introduced urns A and B containing 100 black and red balls each. Urn B is known to have exactly 50 black to 50 red balls. The experimenter asks the subject to choose a color and the urn to pick a ball from. If the chosen color (designated colour) and the drawn color match, a prize of $100 is paid, otherwise the payoff is zero. Instead, we ask the subject first to select the urn if the designated colour is red, and then to select the urn for the black colour. From this point on we refer to this experiment as a standard Ellsberg experiment (task) unless otherwise specified.

The Ellsberg task defines two states, red or black, and a payoff of c_R if a red ball is drawn, otherwise zero. We are therefore in the "gains" domain of the value function and can disregard the reference point. Drop the index by denoting μ the probability of drawing red, and $1-\mu$ the probability of drawing black.

Remark 6.1 *for cumulative prospect theory* $1-\mu$ *is the cumulative probability* $\Pr(c_s \le c_B = 0)$ *of obtaining the worst outcome and* $\Pr(c_s \le c_R) = 1$ *is the cumulative probability of obtaining either the worst or the best. Weightings change to* $\psi(1-\mu)$ *and* $(\psi(1) - \psi(1-\mu))$ *correspondingly.*

The decision maker prefers urn A if and only if

$$\psi(\mu)\, v(\mathrm{R}) + \psi(1-\mu)\, v(\mathrm{B}) > \psi\left(\frac{1}{2}\right) v(\mathrm{R}) + \psi\left(\frac{1}{2}\right) v(\mathrm{B}).$$

When the prize is contingent on the opposite ball colour, the decision rule flips:

$$\psi(\mu)\, v(\mathrm{R}) + \psi(1-\mu)\, v(\mathrm{B}) > \psi\left(\frac{1}{2}\right) v(\mathrm{R}) + \psi\left(\frac{1}{2}\right) v(\mathrm{B}).$$

Substitute for $v(\mathrm{R}) = 1$ and $v(\mathrm{B}) = 0$ in the upper and $v(\mathrm{R}) = 0$ and $v(\mathrm{B}) = 1$ in the lower to obtain

$$\psi(\mu) > \psi\left(\frac{1}{2}\right) \text{ and } \psi(1-\mu) > \psi\left(\frac{1}{2}\right),$$

which implies that for AL-subjects holds

$$\psi(\mu) + \psi(1-\mu) > 2\psi\left(\frac{1}{2}\right).$$

Similarly for AA-subjects, yielding that the weighting function ψ captures am-

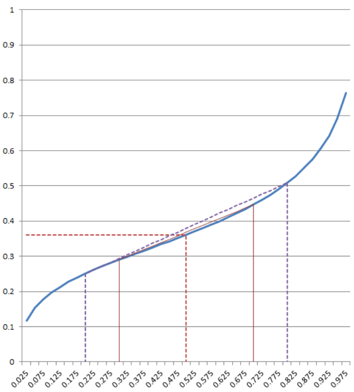

Figure 19. Prospect Theory implicitly assumes ambiguity-loving behaviour.

biguity attitude:

$$\text{AA:} \quad \frac{\psi(\mu)+\psi(1-\mu)}{2} < \psi\left(\frac{1}{2}\right),$$
$$\text{AL:} \quad \frac{\psi(\mu)+\psi(1-\mu)}{2} > \psi\left(\frac{1}{2}\right).$$

For a special case $\psi(\mu) = \frac{\mu^\alpha}{(\mu^\alpha+(1-\mu)^\alpha)^{\frac{1}{\alpha}}}$, which is the widely used parametrization suggested originally by Kahneman and Tversky, one readily obtains $\frac{\psi(\mu)+\psi(1-\mu)}{2} > \left(\frac{1}{2}\right)^\alpha$ and $\psi\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^{\alpha+\frac{1}{\alpha}-1}$. As a result, for any values of μ (without questionning where they come from) and any level of parameter $\alpha < 1$ decision-makers exhibit ambiguity-loving behaviour:

$$\left(\frac{1}{2}\right)^\alpha > \left(\frac{1}{2}\right)^{\alpha+\frac{1}{\alpha}-1}.$$

An illustration of this is in Fig.19.

For cumulative PT:

The decision maker prefers urn A if and only if

$$(\psi(1) - \psi(1-\mu))\, v(\mathrm{R}) + \psi(1-\mu)\, v(\mathrm{B}) > \left(\psi(1) - \psi\left(\frac{1}{2}\right)\right) v(\mathrm{R}) + \psi\left(\frac{1}{2}\right) v(\mathrm{B}).$$

with $v(\mathrm{R}) = 1$ and $v(\mathrm{B}) = 0$ one obtains

$$\psi(1) - \psi(1-\mu) > \psi(1) - \psi\left(\frac{1}{2}\right) \text{ and hence } \psi(1-\mu) < \psi\left(\frac{1}{2}\right).$$

Assuming $v(\mathrm{R}) = 1$ and $v(\mathrm{B}) = 0$ is not necessary; it suffices to know that $v(\mathrm{R}) > v(\mathrm{B})$:

$$
\begin{aligned}
-\psi(1-\mu)\,v(\mathrm{R})+\psi(1-\mu)\,v(\mathrm{B}) &> -\psi\left(\frac{1}{2}\right)v(\mathrm{R})+\psi\left(\frac{1}{2}\right)v(\mathrm{B}). \\
\psi(1-\mu)\,(v(\mathrm{R})-v(\mathrm{B})) &< \psi\left(\frac{1}{2}\right)(v(\mathrm{R})-v(\mathrm{B})) \\
\psi(1-\mu) &< \psi\left(\frac{1}{2}\right) \\
\text{iff } \mu &> \frac{1}{2} \text{ for increasing } \psi.
\end{aligned}
$$

When the prize is contingent on the opposite ball colour ($v(\mathrm{R}) < v(\mathrm{B})$), the decision rule yields the choice of urn A iff $\psi(1-\mu) > \psi\left(\frac{1}{2}\right)$. On the one hand, this demonstrates that cumulative prospect theory is incompatible with Ellsberg paradox. This holds for any mapping ψ, not necessarily a cumulative probability distribution. In other words, in order to make cumulative prospect theory compatible with decisions in ambiguity, decision weights ψ have to be unrelated to possible underlying probabilities μ and meet $\psi(\text{B in urn A}) < \psi(\text{B in urn B})$ and $\psi(\text{R in urn A}) < \psi(\text{R in urn B})$. In this case, ψ captures ambiguity aversion, yet it is rather a capacity (in the sense of Schmeidler, 1989) and not a CDF (or any other mapping in rank-dependent utilities).

Appendix B. Capacities as probability weighting

If ψ is a capacity (nothing prevents it from being one), equation (3) yields some already known rules of choice. Assume ψ is NEO-additive capacity (featuring some other choice criteria as special cases.) The application of the neo-additive capacity (and capacities in general) to second order beliefs should be straightforward as expected utilities are real numbers, and "states" are different probability distributions π. Equation (3) turns into

$$
V(f)=\alpha\int_{\Delta}\phi\left(\int_S u(f)\,d\pi\right)d\mu+(1-\alpha)\left[\delta\min_{\pi}\phi\left(\int_S u(f)\,d\pi\right)+(1-\delta)\max_{\pi}\phi\left(\int_S u(f)\,d\pi\right)\right]. \tag{B-1}
$$

If $\phi(x)=x$ then the first term represents the mean (as given by μ) expected utility, and the two terms in the square brackets represent the lowest and the highest expected utility on the set of "feasible" probability distributions π.

Applied to (2), neo-additive capacities yield

$$V(f)=\int_{\Delta}\phi\left(\alpha\int_{S}u(f)\,d\pi+(1-\alpha)\left[\delta\min_{f}u(f)+(1-\delta)\max_{f}u(f)\right]\right)d\mu, \quad \text{(B-2)}$$

which higlights the above discussion on the role of distortions: the decision-maker here takes an expected value of his/her distorted beliefs BUT does not apply any sort of distortion on μ. If $\phi(x)=x$ the last equation transforms into

$$V(f)=\alpha\int_{\Delta}\int_{S}u(f)\,d\pi d\mu+(1-\alpha)\left[\delta\int_{\Delta}\min_{f}u(f)\,d\mu+(1-\delta)\int_{\Delta}\max_{f}u(f)\,d\mu\right], \quad \text{(B-3)}$$

with the first term, again, corresponding to the mean expected utility and the term in square brackets capturing expected lowest and highest values of utility, contrast to lowest and highest extected utilities. It seems more plausible that a decision-maker rather deviates towards the minimum (maximum) expected value that (s)he can obtain on all feasible probability distributions (B-1) than towards the *average minimum* (maximum) value (as in B-3).

The decision rule (3) stresses that probability weighting is only applicable to ambiguous events. Decisions with regards to unambiguous events, e.g. determined by a toss of a fair coin, are made within vN-M expected utility paradigm. Yet, subjects exhibit different ambiguity attitudes, and even ambiguity neutral subjects have to make decisions when facing ambiguity. These decisions cannot be represented by a maximization of the expected utility functional (as it only refers to unambiguous events) and hence are to be captured by the probability weighting function.

In terms of NEO-additive capacities, ambiguity neutral subject can be represented as one with $\alpha=1$. Applied to unambiguous decisions, deviations from $\alpha=1$ result in "optimistic" and "pessimistic" accounts of the best and worst outcomes. Applied to ambiguous decisions, they result in the corrections for the best and worst expected utilities, as in (B-1), which for ambiguity neutral subjects transforms into

$$V(f)=\int_{\Delta}\phi\left(\int_{S}u(f)\,d\pi\right)d\mu. \quad \text{(B-4)}$$

Effectively, the last equation means that ambiguity-neutral subjects have probability weights $\psi(\mu)=\mu$. This suggests that probability weighting is rather an ambiguity-attitude-driven phenomenon.

For an experimental study of probability weighting functions we use a mechanism (standard Ellsberg two-urn task: ambiguous urn has 100 balls of two colours, unambiguous urn contains 50 balls of each of the two colours, the subject indicates the urn from which a ball should be drawn, and obtains a prize if the drawn ball is of the requested colour.) to select ambiguity-averse, ambiguity-loving and ambiguity-neutral subjects. Further, we elicit probabilities that each of the groups assign to the possible probability distributions in the ambiguous urn (effectively possible distributions of the balls in the urn). Finally, by using the fact that the ambiguity-neutral group assigns probability weights $\psi(\mu) = \mu$, we compare the probabilities assigned by the two other groups with those assigned by the ambiguity-neutral one to elicit probability weightings of ambiguity-averse and ambiguity-neutral subjects.