Discriminating between Models of Ambiguity Attitude: A Qualitative Test

Robin Cubitt[*], Gijs van de Kuilen[+] and Sujoy Mukerji[#]

Forthcoming in the *Journal of the European Economic Association*
19 April 2018

Abstract:

During recent decades, many new models have emerged in pure and applied economic theory according to which agents' choices may be sensitive to ambiguity in the uncertainty that faces them. The exchange between Epstein (2010) and Klibanoff *et al*. (2012) identified a notable behavioral issue that distinguishes sharply between two classes of models of ambiguity sensitivity that are importantly different. The two classes are exemplified by the α-MEU model and the smooth ambiguity model, respectively; and the issue is whether or not a desire to hedge independently resolving ambiguities contributes to an ambiguity averse agent's preference for a randomized act. Building on this insight, we implement an experiment whose design provides a qualitative test that discriminates between the two classes of models. Among subjects identified as ambiguity sensitive, we find greater support for the class exemplified by the smooth ambiguity model; the relative support is stronger among subjects identified as ambiguity averse. This finding has implications for applications which rely on specific models of ambiguity preference.

KEYWORDS: Ambiguity sensitivity; ambiguity attitude; testing models of ambiguity sensitive preference.

JEL-CLASSIFICATION: C91, D01, D03, D81, G02.

ACKNOWLEDGEMENTS: We thank Peter Klibanoff and Peter Wakker for encouragement and detailed comments, as well as an editor, four referees and the audiences and discussants at the various seminars and conferences where we have presented the paper for their many comments. The research was partly funded by a generous gift from an anonymous alumnus of University College, Oxford. Robin Cubitt acknowledges the support of the Economic and Social Research Council (awards ES/K002201/1 and ES/P008976/1) for the Network for Integrated Behavioural Science).

[*] School of Economics and Centre for Decision Research and Experimental Economics, University of Nottingham, Nottingham, NG7 2RD, UK, robin.cubitt@nottingham.ac.uk
[+] Dept. of Economics, Tilburg University, Tilburg, the Netherlands, g.v.d.kuilen@uvt.nl
[#] School of Economics and Finance, Queen Mary, University of London, London, E1 4NS, UK. s.mukerji@qmul.ac.uk

1. Introduction

Decision makers choosing between acts are said to face ambiguity if they are uncertain about the probability distribution over states of the world. Over the past three decades a large decision theoretic literature has developed, inspired partly by the intuitive view that it is often implausible that a decision maker can confidently select a single probability distribution over states of the world to summarize her uncertainty, so ambiguity is ubiquitous for decision making in the real world. This literature, reviewed for example by Etner *et al.* (2012) and Gilboa and Marinacci (2013), also draws important inspiration from numerous experimental studies, largely built on Ellsberg (1961)'s classic examples, which show that subjects often adjust their behavior in response to ambiguity in ways that cannot be accounted for by subjective expected utility theory (for surveys, see Camerer and Weber, 1992; Wakker, 2010; Trautmann and van de Kuilen, 2016). For instance, many subjects display an ambiguity averse attitude: intuitively put, being inclined to choose actions whose consequences are more robust to the perceived ambiguity. Recent applied economic theory explores how such departures from subjective expected utility theory in the face of plausible forms of uncertainty may affect a range of economic phenomena.[1]

The pioneering models in the decision theory literature on ambiguity, and arguably still the most popular, are the Choquet expected utility model of uncertainty aversion introduced in Schmeidler (1989) and the maxmin expected utility (MEU) model of Gilboa and Schmeidler (1989). These models have preference representations which show the decision maker (DM) behaving as if she has a *set* of probability distributions that she considers possible or relevant. Then, an ambiguity averse attitude is modeled by having the DM evaluate an act by its minimum expected utility, where the minimum is taken over the set of probability measures considered possible. In a more general version of this classic style of model (α-MEU; Hurwicz, 1951; Jaffray, 1989; Ghirardato *et al*., 2004), the DM evaluates acts by considering a weighted average of minimum and maximum expected utility. More recent theories have brought in preference representations which would allow finer nuances of ambiguity attitude. An important feature that distinguishes the newer vintage models from the earlier ones is that the new models use aggregation rules that do not restrict attention to extreme expected utilities. An example is the smooth ambiguity model of Klibanoff *et al.* (2005) (KMM).

Given this theoretical development a natural question is: Are the features that these newer theories build in empirically compelling? Or, if we were to stick to the classic models of ambiguity averse behavior, would we miss any empirically important aspect of such behavior? As models in both vintages were designed to capture Ellsberg's classic examples, the many previous experiments based on decisions like those examples do not answer these questions as they do not typically discriminate between the classic and new vintage models. In this paper, we report an experimental study that does

[1] We give examples below. For wider coverage, see for example the discussion and references in Etner *et al.* (2012, Section 7), Gilboa and Marinacci (2013, Section 6) and Trautmann and van de Kuilen (2016, penultimate section).

so discriminate: the two classes of models predict qualitatively different behavior in our design. Thus, the design discriminates between the MEU/α-MEU family of models and the smooth ambiguity model, arguably the most popular models in applications. As we explain in Section 2.3, this divide is not addressed well in existing experimental literature.

This is important since, as noted above, there have been many recent applications of models of ambiguity-sensitive preferences to the understanding of economic phenomena, especially in macroeconomics and financial economics. The typical paper uses a particular preference model, say the MEU model, to explain a phenomenon that is hard to explain plausibly using the standard, expected utility, model. However, some of the explanations depend quite crucially on the particular model of ambiguity-sensitivity used. For example, Epstein and Schneider (2010) discuss various applications where MEU works to give the desired result, but the smooth model does not because it does not generate kinked indifference curves, as MEU does. On the other hand, some recent applications in the macro-finance area, such as Ju and Miao (2012), Jahan-Parvar and Liu (2014), and Collard *et al*. (2018), rely on being able to calibrate beliefs and an ambiguity attitude parameter separately, something which can be done in the smooth ambiguity model, but not in the MEU model. Models of ambiguity averse preference have now also been applied outside macroeconomics and finance, for instance to climate change policy, where similar issues apply (e.g., Millner *et al*. (2013)'s use of the smooth ambiguity model and Chambers and Melkonyan (2017)'s use of MEU). Here too, there is no guarantee that results that hold under one model of ambiguity aversion generalize to other models. The literature is therefore at a point where clearer guidance on the relative empirical performance of these models in particular – and the broader classes that they exemplify – is needed.

Our testing strategy is inspired by the second thought experiment of Epstein (2010) and its generalization in Klibanoff *et al.* (2012). Our main contribution is to recast the generalized thought experiment as a real operational design, to extend it with additional controls, and to run and report the resulting experiment. The testing strategy is to investigate whether a subject's preference for a randomized act (compared to its pure constituents) is influenced by a desire to hedge across ambiguities in a way that is similar to how diversifying across bets on independent risks hedges those risks. Models of preferences whose representations focus exclusively on minimum and/or maximum expected utilities in the set considered relevant are uninfluenced by such a desire, in sharp contrast to models whose representations also consider non-extreme expected utilities. Intuitively, a DM focusing only on minimum expected utility is analogous to an *infinitely* risk averse agent caring exclusively about the worst possible outcome and so not about diversifying across independently resolving risks, since such diversification does not affect the possibility of the worst outcome.

For concreteness and to allow the reader to relate easily to the discussions in Epstein (2010) and Klibanoff *et al.* (2012), we explain our design and results in the main text in terms of the α-MEU and smooth ambiguity models, the divide between which is particularly clear. Appendix C substantiates

our claim that the predictions we test also mark a divide between broader classes of models, besides these two. If (as suggested above) ambiguity is ubiquitous in real world decision making, the importance attached by economists to hedging as a response to uncertainty provides an additional general motivation for our study which goes beyond models, namely to investigate hedging of ambiguities rather than risks.

The rest of the paper is organized as follows. Section 2.1 describes the α-MEU and smooth ambiguity preference representations; Section 2.2 presents a modified version of Epstein's example and uses it to explain our testing strategy; Section 2.3 contrasts this strategy with others taken in the literature. Section 3 presents the experimental design, and Section 4 the results, of our main study. Section 5 introduces some issues of robustness and generality which are examined further in Appendices A, B and C; it also briefly presents a follow-up study in which one aspect of the design of the main experiment is varied, for reasons explained at that point. Section 6 concludes the main text. Appendices D and E contain further details of the results, experimental procedures and instructions.

2. Background

2.1 Preference representations

Formally, the DM's choices are acts, maps from contingent states of the world to consequences, which include simple lotteries with real outcomes. We focus on two models of preferences over acts: the α-MEU model and the smooth ambiguity model. Each captures the idea that the DM does not know the probabilities of the states by postulating a *set* of probability measures over the states that she considers possible. The models differ in respect of how that set informs her evaluations of acts.

In the α-MEU model, an act f is evaluated by:

$$V_M(f) = \alpha \min_{p \in \mathcal{P}} \left(E_p\big(u(f)\big) \right) + (1-\alpha) \max_{p \in \mathcal{P}} \left(E_p\big(u(f)\big) \right),$$

where u is a von Neumann-Morgenstern utility function, α a fixed weight, and $\mathcal{P}$ the set of probability measures p over the states. The operator E_p takes expectations with respect to the measure p. Attitude towards pure risk is characterized by the shape of u, while attitude towards ambiguity is characterized by the weight α; in particular, the greater the value of α, the more ambiguity averse the preference. With $\alpha = 1$, we get the MEU representation.

In the smooth ambiguity model, an act f is evaluated by:

$$V_S(f) = E_\mu \phi \left(E_{p \in \mathcal{P}}\big(u(f)\big) \right),$$

where ϕ is an increasing function mapping utilities to reals, and μ is a subjective probability over the elements of $\mathcal{P}$. The operators E_μ and E_p take expectations with respect to the measures μ and p, respectively. Thus, μ represents the DM's subjective uncertainty about the different probabilities deemed possible and, in this sense, is a second-order belief. Attitudes towards ambiguity are

characterized by the shape of ϕ, given u. In particular, a concave ϕ characterizes ambiguity aversion which is therefore modelled as an aversion to mean preserving spreads in the distribution of expected utilities induced jointly by μ and u. When ϕ is linear or μ is degenerate, the smooth ambiguity model collapses to a subjective expected utility (SEU) model.

Although these models have some common features, there are also marked differences between them, one of which drives our testing strategy as we now explain.

2.2 Conceptual background

Consider the following variant of the second thought experiment proposed in Epstein (2010).[2] The DM is told that a ball will be drawn from an urn containing a fixed number of balls, of 4 different types: B_1, B_2, R_1, R_2. She is also told that the combined number of balls of types B_1 and B_2 will equal that of balls of types R_1 and R_2; and, finally but importantly, that the relative proportions *within* the *B-component* (B_1, B_2) and within the *R-component* (R_1, R_2) will be determined separately. The DM considers acts with contingent outcomes c, c^* and the 50-50 lottery between them. Let $c^* > c$ and normalize the utility index u, so that $u(c^*) = 1$ and $u(c) = 0$. The acts to be considered have state-contingent (expected) utility payoffs as described in Table 1.

To clarify, f_1 yields c^* when a ball of type B_1 is drawn and c otherwise; whereas f_2 yields c^* when type R_1 is drawn and c otherwise. The outcome of the act *mix* is in part decided by the toss of a fair coin: specifically, for any contingency, there is a 0.5 probability that the outcome is determined by applying f_1 and a 0.5 probability that it is determined by applying f_2. Below, "mixed act" always refers to this mixed act and "constituent acts" to f_1 and f_2 (or, in each case, later to their experimental counterparts). The acts g_1 and g_2 each yield, in the contingencies for which a cell-entry of ½ is shown, either c^* or c, depending on the toss of a fair coin (and c otherwise).

Table 1: Five Acts: (Expected) utilities

	B_1	B_2	R_1	R_2
f_1	1	0	0	0
f_2	0	0	1	0
mix	$\frac{1}{2}$	0	$\frac{1}{2}$	0
g_1	$\frac{1}{2}$	$\frac{1}{2}$	0	0
g_2	0	$\frac{1}{2}$	$\frac{1}{2}$	0

How might we expect the DM to choose between these acts? The probability of the event $\{B_1, B_2\}$ is objectively known to her and equal to ½ (as types B_1 and B_2 jointly account for half of the balls), but

[2] Readers familiar with Epstein (2010) and Klibanoff *et al.* (2012) should note that our variant differs in some respects from Epstein's thought experiment and that we adopt a notation that is suited to our experimental design in ways that will emerge below but does not match the earlier papers.

the DM does not know the probability of the event $\{B_2, R_1\}$. Moreover, the information that she has about balls of type B_1 exactly matches her information about type R_1. Thus, the symmetry in the situation suggests $f_1 \sim f_2$; and, it is natural to expect that, if the DM is ambiguity averse, she will have the strict preference $g_1 \succ g_2$.[3]

While there may be little to disagree about in these claims, it is much more controversial whether an ambiguity averse DM would see mixing f_1 and f_2 as desirable, compared with either of the latter two acts alone. This issue illustrates one of the main points of contention between the two models considered in the previous subsection. The issue is whether, for an ambiguity averse DM who is indifferent between f_1 and f_2,

$$f_1 \sim f_2 \sim mix \quad \text{or} \quad f_1 \sim f_2 \prec mix$$

would obtain. In the α-MEU model, the former condition will hold whereas, in the smooth ambiguity model, the latter condition will hold. In words, the issue which divides the models is whether, in this situation, an ambiguity averse DM would or would not value the averaging that the mixed act offers.

To illustrate the point of contention, it is useful to write down a concrete set of probability measures $\{p_1, \ldots, p_4\}$ that we suppose to be those considered by the DM. In the context of the smooth ambiguity model, think of these as probabilities that are given positive weight by the measure μ and, importantly, with the weights for p_2 and p_3 equal.

Table 2: Example Probabilities

	B_1	B_2	R_1	R_2
p_1	$\frac{1}{10}$	$\frac{4}{10}$	$\frac{1}{10}$	$\frac{4}{10}$
p_2	$\frac{1}{10}$	$\frac{4}{10}$	$\frac{4}{10}$	$\frac{1}{10}$
p_3	$\frac{4}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{4}{10}$
p_4	$\frac{4}{10}$	$\frac{1}{10}$	$\frac{4}{10}$	$\frac{1}{10}$

These measures respect the given information, in that, for each $i = 1, 2, 3, 4$, $p_i(B_1 \cup B_2) = p_i(R_1 \cup R_2) = ½$; and, as p_2 and p_3 have equal weight, there is complete symmetry between the B-component and the R-component. The measures respect independence of the two components in that fixing a "marginal" over (B_1, B_2) does not restrict the "marginal" over (R_1, R_2), or vice versa. The expected utilities generated by applying each of the measures p_i from Table 2 to the acts from Table 1 are as follows:

[3] Correspondingly, if $f_1 \sim f_2$ and the DM is ambiguity seeking, one would expect $g_2 \succ g_1$. Ambiguity neutrality as represented in the model of Anscombe and Aumann (1963) implies that the direction of preference between f_1 and f_2 should match that between g_1 and g_2, so that $f_1 \sim f_2$ would imply $g_1 \sim g_2$.

Table 3: Resulting Expected Utilities

	p_1	p_2	p_3	p_4
f_1	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{4}{10}$	$\frac{4}{10}$
f_2	$\frac{1}{10}$	$\frac{4}{10}$	$\frac{1}{10}$	$\frac{4}{10}$
mix	$\frac{1}{10}$	$\frac{2.5}{10}$	$\frac{2.5}{10}$	$\frac{4}{10}$
g_1	$\frac{2.5}{10}$	$\frac{2.5}{10}$	$\frac{2.5}{10}$	$\frac{2.5}{10}$
g_2	$\frac{2.5}{10}$	$\frac{4}{10}$	$\frac{1}{10}$	$\frac{2.5}{10}$

First, consider acts f_1 and f_2. Their expected utilities coincide under p_1 and p_4, but differ from each other under p_2 and p_3. To see why, note from Table 1 that the evaluation of f_1 depends on the ratio B_1:B_2 but not on R_1:R_2; whereas the evaluation of f_2 depends on the ratio R_1:R_2 but not on B_1:B_2; and then note, from Table 2, that these ratios coincide under p_1 and p_4, but not under p_2 or p_3. In contrast, the evaluation of *mix* depends on both the ratios, but has half the exposure to the uncertainty about each, compared to each of the constituent acts. The point of contention turns on the significance of these facts.

From the perspective of the α-MEU model, the extremes of the possible expected utilities are what matters for the evaluation of an act. The diversification aspect of the comparison between f_1, f_2 and *mix* is irrelevant, as the minimum and maximum possible expected utilities are the same under each of these three acts, as Table 3 shows. So, according to this model, the DM will be indifferent between f_1, f_2 and *mix*, regardless of her preference over g_1 and g_2.

However, from the perspective of the smooth ambiguity model, the mixed act provides a hedging of two separate ambiguities, one involving each of the two components, just as diversifying across bets on independent risks provides a hedging of risks. The benefit of such diversification to an ambiguity averse DM is captured through a concave ϕ, in that mean-preserving spreads in the subjective distribution of expected utilities generated by an act are disliked. Since p_2 and p_3 have equal weight, each of f_1 and f_2 yields a mean-preserving spread in expected utilities compared with *mix*, as Table 3 shows. Thus, according to the smooth ambiguity model, the mixed act is preferred to its constituents by any ambiguity averse DM. To generalize, the distinctive prediction of the smooth ambiguity model for the case where p_2 and p_3 have equal weight is that an ambiguity averse DM will prefer not just g_1 to g_2 but also *mix* to each of f_1 and f_2; and, correspondingly, an ambiguity seeking DM (convex ϕ) would have the reverse preference in each case; and an ambiguity neutral DM (linear ϕ) would be indifferent between g_1 and g_2, and between *mix* and each of its constituents.

It is important to note a key feature of the perspective of the smooth ambiguity model. Each of p_2 and p_3, the measures across which the mixed act smooths expected utility relative to its constituents, corresponds to a situation where there is one "marginal" over component (B_1, B_2) and a different "marginal" over (R_1, R_2). Thus, it is precisely because it is uncertain whether the two components are identical to one another (so leading the DM to consider p_2 and p_3) that the diversification provided by the mixed act is seen by the smooth model as valuable to an ambiguity averse DM. If, instead, the two components were known to be identical (and so only p_1 and p_4 considered), smooth ambiguity

preferences would display indifference between the mixed act and its constituents, just as α-MEU preferences would. Thus, the key difference between smooth ambiguity preferences and α-MEU preferences that we have highlighted is whether the DM values hedging across ambiguities that are separate, in the sense that the uncertainty about the probability governing one component resolves separately from the analogous uncertainty for the other component. This insight is crucial to our experimental design, as explained in Section 3.

2.3 Related literature

Our experimental design identifies subjects whose behavior is sensitive to ambiguity, categorizing them as ambiguity averse or seeking, and determines whether they behave according to the α-MEU model or the smooth ambiguity model in a set-up of the kind described in Section 2.2. The tests of whether this is so rely on qualitative features of the data, i.e., binary preferences (revealed, as explained below, by ordinal comparisons of certainty equivalents). None of our tests require estimates of model parameters. It is useful to bear these points in mind as we discuss how this experiment fits in with other recent literature. We concentrate on papers whose main objective is to distinguish empirically between models similar to those we consider.[4]

The experimental approach of Halevy (2007) is to determine whether a subject may be classified as ambiguity neutral/averse/seeking (using an Ellsberg style determination), while also checking how the subject evaluates an *objective* two-stage lottery, in particular whether the evaluation is consistent with reduction of objective compound lotteries (ROCL).[5] The main finding is that ambiguity aversion is strongly associated with violation of ROCL. Using this finding, the study sifts evidence for or against various models of ambiguity sensitivity. For instance, while the α-MEU model predicts a zero association with ROCL, in several models in (what Halevy terms) the "recursive expected utility" class, ambiguity sensitivity logically implies violation of ROCL. However, under the assumptions of KMM, there is no logical connection between ambiguity aversion (or, seeking) and reduction of compound *objective* lotteries in the smooth ambiguity model.[6] Hence, the strategy based on ROCL is not as useful in distinguishing α-MEU from the smooth model in KMM as it is in making other distinctions.[7]

---

[4] Attanasi *et al*. (2014) tests the smooth model using its predictions in a portfolio choice experiment, but does not aim to discriminate between models. Many other empirical papers are primarily concerned to measure ambiguity sensitivity rather than to test models. Cubitt *et al*. (2018) does this using the smooth ambiguity model and Ahn *et al*. (2014) using both smooth ambiguity and α-MEU models. Trautmann and van de Kuilen (2016) surveys a range of model-free measures of the sensitivity.

[5] See Abdellaoui *et al*. (2015) for an extension of this testing strategy.

[6] See the discussions following KMM's Remark 1 and immediately preceding and in footnote 8 in Klibanoff *et al.* (2012).

[7] It is important that we refer to the smooth model as developed by KMM here. Seo (2009) obtains a representation that uses an identical functional form to represent preferences defined on a choice domain that differs from KMM's by including multi-stage lotteries, but excluding second-order acts. In this domain and under the axioms applied by Seo (but unlike in KMM), violation of ROCL *is* equivalent to a concave ϕ. Because of this, Halevy's experiment *does* distinguish between MEU, SEU and Seo's preference model. In Section 5.2, we report on a

Conte and Hey (2013) observe subjects' choices between prospects and study how well the data fit various models of decision making. Unlike that of Halevy, the identification strategy is not based primarily on qualitative features of the data. Instead, they estimate parametric preference models, in particular, the SEU, α-MEU and smooth ambiguity models. One part of the study fits the models subject-by-subject, while another part estimates a mixture model. However, the uncertain prospects the subjects are given to choose between are still objective two-stage lotteries of the kind used in Halevy (2007). So, the point still applies that subjects who are strictly ambiguity averse/seeking, and whose preferences conform to the smooth model, may not evaluate such lotteries any differently from those whose preferences satisfy expected utility theory. Hey *et al*. (2010)[8] and Hey and Pace (2014) also compare the descriptive and predictive performance of particular parameterisations of several "non-two-stage probability models" of behaviour, in these cases using ambiguity that is generated by a bingo-blower. But the smooth ambiguity model is not one of those they consider and, despite their attractiveness in some contexts, it is unlikely that bingo-blower designs could deliver the control over beliefs that our design exploits, as we explain in the next section.

Taking a different approach, Ahn *et al.* (2014)'s experiment studies a simulation of a standard economic choice problem: each subject allocates a given budget between three assets, each of which pays depending on the color of the ball drawn from a single Ellsberg style three color urn, while the prices of assets are exogenously varied.[9] Different parametric preference models of choice under uncertainty imply different asset demand functions. Ahn *et al.* (2014)'s testing strategy distinguishes quite effectively between two classes of models: those that have kinked indifference curves (e.g., α-MEU and the rank dependent utility model) and those with smooth indifference curves (e.g., SEU and smooth ambiguity, even if ambiguity averse), as kinked and smooth indifference curves imply demand functions with different qualitative characteristics in their setting. However, the identification is more problematic *within* each class of model. Indeed, if a subject's preferences are ambiguity averse and conform to the smooth ambiguity model, qualitative properties of choice data in this experiment do not distinguish her from an SEU subject. Similarly, an α-MEU preference is difficult to distinguish qualitatively from first order risk aversion as models of first order risk aversion where preferences are fully probabilistically sophisticated in the sense of Machina and Schmeidler (1992) also bring in kinks in different ways – for example, the rank dependence model of Quiggin (1982) and prospect theory of Tversky and Kahneman (1992).

Hayashi and Wada (2009) investigate choice between lotteries where the subject has imprecise (objective) information about probability distributions defining the lotteries. While they do not

follow-up to our main experiment motivated by the divide between the smooth model of KMM and the model of Seo (2009).

[8] See also Kothiyal *et al*. (2014).

[9] Bossaerts *et al.* (2009)'s experiment employs a choice problem very similar to Ahn *et al.* (2014)'s, but in the context of a market where prices are determined by trading among subjects.

specifically test the α-MEU model against a smooth ambiguity model, a finding relevant to our discussion is that their subjects appear to care about more than just the best-case and worst-case probability distributions. However, their strategy for detecting this influence of non-extreme points (in the set of possible probabilities) does not exploit the hedging motive that we stress.

In contrast, Andreoni *et al*. (2014) studies subjects' attitudes to mixtures between subjective and objective bets. As different models that allow for ambiguity sensitivity relax the Independence axiom in different ways, their test can potentially separate some of the different models (in particular MEU vs SEU vs the smooth ambiguity model). But, they point out that their test of the smooth ambiguity model is conditional on particular functional forms, unlike the one we apply in the present study.

Finally, Baillon and Bleichrodt (2015) use elicited "matching probabilities" to distinguish between several models, including the α-MEU and smooth ambiguity models. Their approach differs from ours by its use of indifferences expressed on a probability scale, rather than on a monetary scale, to indicate preferences; and by their focus on the ability of models to account for observed differences in ambiguity attitudes between the domains of gains and losses respectively.[10] We set the latter issue aside by concentrating on the domain of gains, in order to focus on the hedging issue at the heart of the dispute between Epstein (2010) and Klibanoff *et al.* (2012) to which we now return.

3. Experimental Design

3.1 *Core of design*

Our design has at its heart an implementation of the theoretical set-up of Section 2.2. In place of an ambiguous urn containing balls of four different types, we used specially constructed decks of cards, divisible into the four standard suits. We implemented the component (B_1, B_2) as the composition by suit of the black-suit (henceforth black) cards and the component (R_1, R_2) as the composition by suit of the red-suit (henceforth red) cards, specifically B_1 = spade, B_2 = club, R_1 = heart and R_2 = diamond. Subjects were told that there would be equal numbers of black and red cards in each deck, but not exactly how the black cards would subdivide into clubs and spades, nor how the red cards would subdivide into hearts and diamonds.

A key feature of our design is that we manipulated whether the compositions of black cards and red cards were mutually dependent or mutually independent. In each case, the compositions were determined by drawing from a bag containing two types of balls, the relative proportions of which were unknown to subjects. In our '1-ball' condition, a single ball was drawn and its type determined the compositions of both the black cards and the red cards, making those compositions mutually dependent. In our '2-ball' condition, two balls were drawn with replacement: the first to determine the composition

[10] This issue is also considered by Dimmock *et al*. (2015) who – like Baillon and Bleichrodt (2015) - use matching probabilities to test various models, in their case including the α-MEU model but not the smooth ambiguity model.

of the black cards, the second to determine the composition of the red cards, making the two compositions mutually independent.

Subjects were informed of these procedures and our analysis uses as an identifying restriction that they believed what they were told. As we explain in Section 3.3, the information given to subjects implied that the set of possible compositions of the whole deck corresponded, in the 1-ball condition, to $\{p_1, p_4\}$ from Table 2 and, in the 2-ball condition, to $\{p_1, p_2, p_3, p_4\}$, with the compositions corresponding to p_2 and p_3 having equal (but unknown) likelihood. Thus, the 2-ball condition implements exactly our variant of the Epstein example, explained in Section 2.2. This allows us to discriminate between the α-MEU and smooth ambiguity preference models using their predictions for that case described earlier. In contrast, because it has no deck compositions corresponding to p_2 and p_3, the 1-ball condition provides a control that eliminates the scope for strict preference for the mixed act over its constituents to derive from the hedging motive postulated by the smooth ambiguity model. If that motive is the only driver of strict preference between the mixed act and its constituents in the 2-ball condition then - and according to both models of Section 2.1 - we would not observe such strict preference in the 1-ball condition. A different possibility is that there are other factors - not captured by either model of Section 2.1 - that give rise to strict preference between the mixed act and its constituents in the 1-ball condition. In this case, we can assess whether the hedging motive postulated by the smooth ambiguity model contributes to preference over the acts by using our two conditions alongside each other, with the 1-ball condition controlling for the role of the other factors.

3.2 Presentation of acts

Acts were presented to subjects as "gambles", the outcomes of which would depend, as just indicated, on the suits of cards drawn from decks. We used two protocols, one verbal and the other tabular, in different sessions, to describe the acts and the construction of the decks to subjects. The results proved insignificantly different and, in Section 4, we pool results from both types of session. Here, we report the tabular protocol in the main text and indicate how the verbal protocol differed from it in footnotes.

In the tabular protocol, acts were described by rows in tables – like Table 4 - of which the column-headings were suits and the cell entries indicated the results, under each given act, of a card of each suit being drawn. The cell entries indicated either that the act would yield €20 in the relevant contingency; or that it would yield €0 in that contingency; or that the outcome in the relevant contingency would depend on a roll of a (standard 6-sided) die in the following way: €20 if the roll was even and €0 if it was odd.[11] Table 4 has a row corresponding to each of the acts from Table 1. Subjects never had to

[11] We used the die-rolling procedure as it is easier to perform reliably in the lab than tossing a coin. With this amendment, suits as contingencies, c^* = €20 and c = €0, the experimental acts exactly match their theoretical counterparts in Table 1.

consider all these acts at once. Instead, they saw tables like Table 4, but with only those rows for the acts they were required to consider at a given point (see below).[12]

Table 4: Description of the Acts

	Spade	Club	Heart	Diamond
f_1	€20	€0	€0	€0
f_2	€0	€0	€20	€0
Mix	Roll die is EVEN: €20 Roll die is ODD: €0	€0	Roll die is EVEN: €20 Roll die is ODD: €0	€0
g_1	Roll die is EVEN: €20 Roll die is ODD: €0	Roll die is EVEN: €20 Roll die is ODD: €0	€0	€0
g_2	€0	Roll die is EVEN: €20 Roll die is ODD: €0	Roll die is EVEN: €20 Roll die is ODD: €0	€0

3.3 Decks

Each act was resolved using one of three 10-card decks that subjects were informed would be constructed after they had completed the experimental tasks. Subjects were also told that, after each deck had been constructed, it would be shuffled and placed face down in a pile. A 10-sided die would then be rolled and the card "drawn" from the deck would be the one whose position in the pile matched the number on the die. These processes were conducted publicly, making it transparent that any card could be drawn from a given deck and that neither the experimenter nor subjects' choices could influence which one was.

At the start of the experiment, subjects completed tasks relating to two risky acts which would be resolved using Deck 1, which subjects were told would contain 7 spades and 3 hearts. These risky acts served as a simple introduction to the experiment for subjects and, as they would be resolved with a deck of known composition, made it more salient that the remaining acts would be resolved using decks about which subjects had only limited information. Those decks (Decks 2 and 3) are our main focus.

As explained in Section 3.1, our design is premised on an assumption that, in each condition (i.e. 1-ball or 2-ball), subjects believed certain sets of compositions of the decks to be those possible. To ground this assumption without compromising the ambiguity of ambiguous acts or deceiving subjects, we employed a strategy with three elements: (i) we used a process to construct the relevant decks that

[12] In the verbal protocol, each act was described by a single line of text, indicating how the outcome would depend on the card drawn from a deck. Subjects were told that, depending on the card drawn, each gamble could have one of three outcomes: WIN, LOSE, or ROLL. WIN and LOSE would yield payments to the subject of €20 and €0, respectively; whereas ROLL would yield €20 if a standard 6-sided die rolled at the end of the experiment came up even, and €0 if it came up odd. Each line of text had the following form: "If _____ you _____, otherwise you LOSE" where the first placeholder was filled either by a single suit (e.g. "Spade") or a disjunction over two suits (e.g. "Spade or Club") and the second placeholder by either WIN or ROLL.

allowed us to control which compositions for each deck were possible in fact; (ii) we told subjects enough about that process to reveal which compositions were possible but not so much as to give objective probabilities over the possibilities; and (iii) we conducted the process publicly at the end of each session.

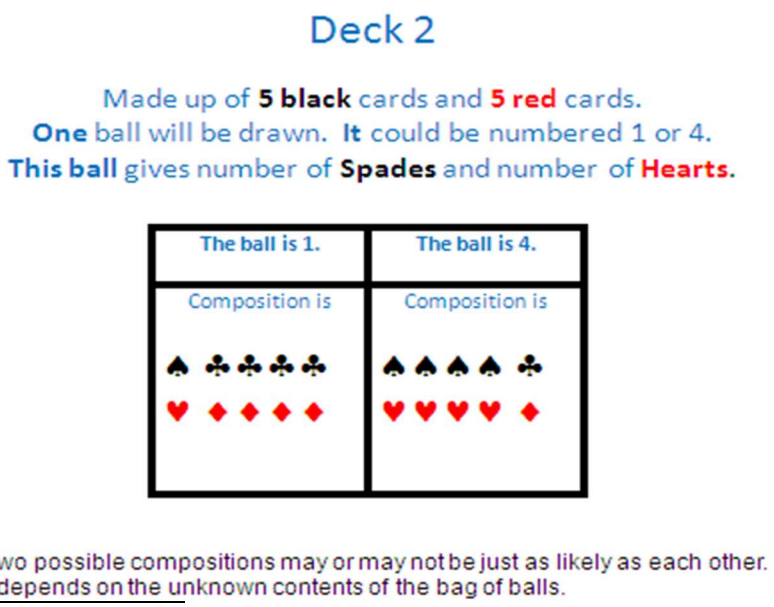
For each of Decks 2 and 3, subjects were told that the deck would consist of 5 black cards and 5 red cards; and, in addition, that the number of spades would be either 4 or 1, with clubs adjusting accordingly; and, similarly, that the number of hearts would be either 4 or 1, with diamonds adjusting accordingly. What subjects were told beyond this varied between Decks 2 and 3, with the different instructions employing in different ways an *opaque bag* containing balls numbered either 1 or 4.

In the *1-ball condition*, tasks concerned acts to be resolved using Deck 2. Before completing these tasks, subjects were told that, at the end of the experiment, one ball would be drawn from the opaque bag. The number on it would give both the number of spades and the number of hearts in Deck 2. Thus, in that deck, the number of spades and the number of hearts would be identical.

In the *2-ball condition*, tasks concerned acts to be resolved using Deck 3. Before completing these tasks, subjects were told that, at the end of the experiment, two balls would be drawn from the opaque bag, *with replacement*. The number on the first ball would give the number of spades in Deck 3 and the number on the second ball would give the number of hearts in Deck 3. Thus, in that deck, the number of spades and the number of hearts would be independent draws.

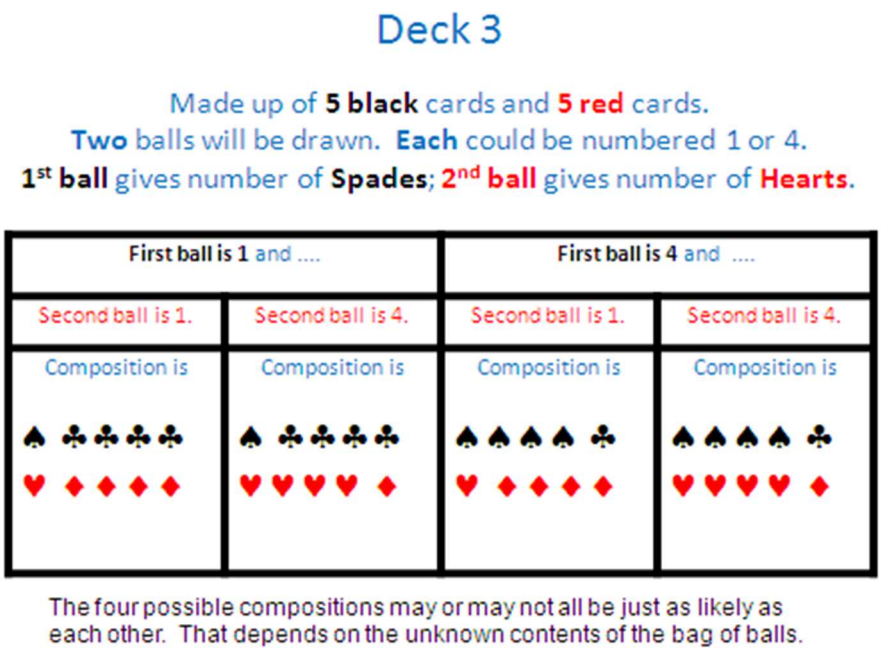
In each condition, the information just specified was conveyed to subjects by projection of slides onto the wall of the lab, while the experimenter described the relevant procedures. The slides for the tabular protocol are as shown in Figures 1 and 2, for the 1-ball and 2-ball conditions respectively.[13]

Figure 1: Deck 2 (1-ball condition)



[13] Each suit (the symbol and corresponding word) appeared in its own color, as did the words black and red, with other text blue. In the verbal protocol, slides described the compositions in words rather than pictorially.

Figure 2: Deck 3 (2-ball condition)



Several features of these procedures are worth stressing. As just explained, in both conditions, subjects were told that the compositions of Decks 2 and 3 would be determined by drawing the appropriate number of balls from the opaque bag, but were *not* told anything more than this about the contents of the opaque bag except that it contained balls of which some were numbered 1 and the others numbered 4. Since subjects did not know the relative frequency of types of ball in the bag, they had no objective probabilities for the possible compositions of either Deck 2 or Deck 3. Thus, for acts resolved with these decks, subjects faced genuine ambiguity, <u>not</u> two-stage objective lotteries. Indeed, the first stage of resolution of uncertainty (i.e. determination of the composition of decks 2 and 3) was ambiguous in just the same way as is a draw from an Ellsbergian urn containing specified types of object in unspecified proportions.[14]

As the process determining the outcome of any given ambiguous act was conducted publicly, subjects were able to verify their limited information about it: balls *were* drawn from the opaque bag as described; they *were* numbered 1 or 4; and each of Decks 2 and 3 *was* constructed to have the composition specified for that deck by the relevant slide, given the draws from the opaque bag. As long as subjects attended to what they were told and realised we would not make statements about procedures

[14] In making explicit mention of a prior stage which determines the probability distribution over payoff relevant contingencies, our design departs from the classic Ellsbergian one. It creates an environment that is closer than the classic Ellsberg case to those considered in many recent applications of ambiguity frameworks, namely, instances of agents facing *model uncertainty*, e.g. in macroeconomics and asset pricing (Marinacci, 2015). In these applications, model uncertainty is precipitated by a suspected regime change, leaving the decision maker uncertain about which stochastic data generating process is in operation in the immediate/near future. The decision maker is typically considered as having specific models of that process in mind, but not having reliable probabilistic information as to which of them yields the "true" distribution.

that would be publicly exposed as false later in the session, we are justified in assuming that subjects believed that each deck would have one of the compositions we had stated to be possible for it.

Some authors, such as Charness *et al.* (2013), conjecture that subjects of previous experiments may have suspected that devices used to resolve ambiguity might be stacked against them by the experimenters. Our design is structured so that, as long as subjects believed information which they would be able to verify, such suspicions would be minimised; and so that any remaining ones would not undermine our objectives. Some subjects may have considered the possibility that, when filling the opaque bag before the session, we might work against their interests but, provided they remained subjectively uncertain of the content of the opaque bag (e.g. because of also considering other scenarios about our characters), the required ambiguity would have remained.[15]

Although subjects had no information about the relative likelihoods of the two possible compositions of Deck 2, nor about those of the first and fourth possible compositions of Deck 3 (relative either to each other or the other two compositions), the information given to subjects *implied* that the second and third possible compositions of Deck 3 were equally likely. (As draws from the opaque bag were with replacement, 1 followed by 4 was precisely as likely as 4 followed by 1.) This is significant in relation to Section 2.2 as it means that a subject who understood the implications of what they were told would attach equal weight to possible compositions of Deck 3 corresponding to p_2 and p_3. Of course, we cannot be sure that all subjects appreciated this. Nevertheless, the information that they were given was entirely symmetric between the second and third possible compositions of Deck 3, so even a subject who did not see that their information implied equal likelihood of those compositions would have been given no grounds for thinking either of them more or less likely than the other. In view of these points, we start from a maintained hypothesis that subjects did weight the second and third possible compositions of Deck 3 equally, as well as believing the possible compositions of each deck to be those we had stated to be possible for it. We discuss the robustness of our conclusions with respect to violations of this maintained hypothesis in Appendix A.

Finally, the 2-ball condition is not just a device for giving subjects a reason to put equal weight on the second and third possible compositions of Deck 3. It also implements the key feature of the theoretical framework of Section 2.2 that the DM understands that the uncertainties about the "*B*-component" (here, the relative frequency of spades and clubs) and uncertainties about the "*R*-component" (here, the relative frequency of hearts and diamonds) are resolved *separately*. This would

[15] Even subjective certainty about the content of the opaque bag would not remove uncertainty about the compositions of decks 2 and 3, unless subjects were subjectively certain that the opaque bag contained only one type of ball. We think that unlikely but, in any case, the design builds in an additional fail-safe. A subject who was subjectively certain that the opaque bag contained only one type of ball would attach weight only to the deck compositions corresponding to p_1, or only to those corresponding to p_4. Either way, they should be indifferent between g_1 and g_2 in the 2-ball condition and so (see below) would be coded as ambiguity neutral and, as a result, would not impinge on our comparisons of models.

not have been achieved by an (arguably more classically Ellsbergian) design in which the four possible compositions of Deck 3 were simply listed for subjects.

3.4 Elicitation of preferences

Our procedure for inferring a preference between two acts was to elicit a certainty-equivalent for each of them and to infer the binary preference from the relative magnitudes of the certainty-equivalents. This procedure allows incentivized elicitation of indifference between two acts while avoiding the problems of choice-tasks in which subjects are allowed to express indifference directly.[16] To infer a subject's certainty-equivalent of a given act, we used a form of *choice-list* procedure that yielded interval estimates with a bandwidth of €0.05. The procedure is similar to that of Tversky and Kahneman (1992), sharing with it the important feature that, because estimated certainty equivalents are obtained from choices, they should be unaffected by endowment effects.

In our case, the details of this procedure were as follows. Acts were displayed to subjects and choice-lists completed by them on computers. The experiment was programmed using z-Tree (Fischbacher, 2007). Each choice-list consisted of a table, each row of which described a choice between an act and a certain sum of money. Comparing successive rows of a given choice-list, the sums of money rose moving down the table, but the act remained the same.[17] In a *basic* list, the first row was a choice between the relevant act and €0; the certain sum of money then rose by €1 per row, till the final row was a choice between the act and €20. (See Appendix E for example basic lists.) As, for each act in our design, the two possible final outcomes were €20 and €0, we obviously expected subjects to choose the act in some early rows (at least the first one); to switch to the certainty in some subsequent row; and then to choose the certainty in all remaining rows. After completing all rows of a basic choice-list to their satisfaction, subjects had to confirm their choices; the computer would only accept confirmed responses with the single-switch property just described (or with no switches). After confirmation of their responses to a basic choice-list, a subject who had switched proceeded to a *zoomed-in* list for the same act. This had the same structure as the basic one, except (i) that the first and last rows were, respectively, the two choices where the subject had switched from the act to the certainty in the basic list, with the responses to these rows filled in as the subject had already confirmed them; and (ii) across the intervening rows the certain sums of money rose in increments of €0.05. Again, the subject was required to choose between the act and each certain sum, observing the single switch

[16] If subjects are presented with the choice between two acts and allowed, as a third option, to express indifference, there would be a problem incentivizing the task. If the third option yields randomization over the other two then arguably, when it is taken, what would have been revealed is strict preference for the randomization, rather than indifference between the initial two acts (unless the nature of the randomization is unknown to subjects – but then there would be a worse confound via an unwanted role for ambiguity attitude).

[17] For support for use of choice lists rather than attempting to elicit an indifference point directly, see Cason and Plott (2014); and, for using choice lists where amounts of money, but not the uncertain option, vary from row to row, see Bosch-Domenech and Silvestre (2013). In respect of the latter point and in using "zooming-in", our design parallels the "iterative multiple price lists" of Andersen *et al*. (2006).

requirement (and could adjust their responses until they confirmed them). A subject's certainty equivalent was coded as the average of the certain sums in last row of the zoomed-in list in which she chose the act and the first row in which she chose the certain sum.

3.5 Incentives

Each subject completed basic lists for ten acts, plus the corresponding zoomed-in lists. They were told at the start that, after they had completed all choices in all choice lists, one such choice would be selected at random to be for real[18]: that is, if they had chosen the certain sum of money in it, they would receive that sum; and, if they had chosen the act, they would receive the outcome of its resolution.[19] This is a form of the *random lottery incentive system*, widely-used in individual choice experiments. It prevents confounding income effects between tasks which might arise if more than one task was paid (likewise, Thaler and Johnson (1990)'s "house money" effects). It is easy for subjects to understand and, in the current context, allows us to elicit certainty-equivalents without using cognitively more demanding devices such as auctions or forms of the Becker-De Groot-Marschak mechanism (Becker *et al*., 1964) in which buying or selling "prices" are declared and compared with randomly drawn ones.[20]

3.6 Sequence of tasks

After the choice-lists for the risky acts to be resolved with Deck 1, subjects completed choice-lists for the ambiguous acts f_1, f_2 and *mix* in the 1-ball condition (Deck 2), followed by choice-lists for the ambiguous acts f_1, f_2, *mix*, g_1 and g_2 in the 2-ball condition (Deck 3). This progression from a risky environment to environments with progressively more complex ambiguity provided a natural sequence, conducive to subjects' understanding.

---

[18] The way in which this worked is significant. Subjects were told that the computer would select at random one row from one basic choice-list. If the task in this row was neither the last in which the subject chose the act nor the first in which she chose the certainty, the subject would receive her choice in the selected row. Otherwise, the computer would select at random a row from the zoomed-in list defined by the subject's choices in the selected basic list, and the subject would receive her choice in that row. This procedure has the important property that the subject's choices have no effect on which row of which basic choice-list is selected by the computer at the first stage; and, if the second stage is reached, no effect on which row of the relevant zoomed-in list is selected. This avoids strategic concerns that would arise if all choices faced were selected to be for real with equal probability.

[19] A card was drawn from each deck at the end of the experiment and a 6-sided die was rolled. Together, these resolved all chosen acts. All sums due were paid in cash before subjects left the experiment.

[20] See Cubitt *et al.* (1998), Bardsley *et al.* (2010, Section 6.5), Baltussen *et al.* (2012), and Azrieli *et al*. (2016) for discussions of the random lottery incentive system. If, as hypothesized theoretically by Holt (1986), Bade (2015) and Baillon *et al.* (2015), there is a tendency for subjects to see all their task-responses as steps in the construction of a single portfolio, that could in principle affect responses to individual tasks. However, it is unclear whether such effects will exist in practice even when there is a theoretical rationale for one, as subjects may take the cognitively simpler route of considering one task at a time. (A typical subject in our design faces 400 binary choices across all choice-lists, so considering a portfolio of all responses would be very demanding.) In the context of risky choice, Starmer and Sugden (1991), Beattie and Loomes (1997), Cubitt *et al.* (1998) and Hey and Lee (2005) found little evidence of any portfolio effect even in cases where Holt (1986)'s theoretical argument suggested there "should" be one. In any case, our design is robust in two further ways. Provided any portfolio effect is the same in the 1-ball and 2-ball conditions, it would not affect the comparison of them. Moreover, as long as the attractiveness of the mixed act is reduced by background randomization, that would tend to militate against, rather than for, it. Given this, the direction of any portfolio effect in either condition can be anticipated.

Our design was constructed to make it "easy" for subjects to express indifference between the acts f_1, f_2 and *mix*. In each condition, all basic choice lists for these three acts were shown and completed side-by-side on the same screen; and subjects then proceeded to the corresponding zoomed-in lists, again with the lists for the three acts side-by-side on the same screen. As subjects could adjust their responses at any time until they confirmed them, they could easily align (or disalign) their certainty equivalents for the acts appearing on the same screen.

After subjects had completed all choice-lists for the mixed act and its constituents in the 1-ball condition and then in the 2-ball condition, they proceeded to a further screen with the basic choice-lists for g_1 and g_2. They were completed side-by-side on the same screen, as were the corresponding zoomed-in lists. As the certainty-equivalents for these acts would be used to categorize subjects by ambiguity attitude (as we explain in the next sub-section), we decided to elicit them last to rule out any possibility that subjects could construct their other choices deliberately to make them consistent with these ones. Subjects did not see these acts at all until after they had completed all tasks involving *mix* and its constituents in the 2-ball condition. Full experimental instructions are given in Appendix E2.

3.7 Classification of subjects

As it was resolved with Deck 3, Table 4 and Figure 2 show that g_1 offered 5 chances (out of 10) of a 50-50 die-roll under every possible composition of the deck. In contrast, g_2 would yield the die-roll if a club or a heart was drawn; and the combined number of clubs and hearts was uncertain. Specifically g_2 offered 5, 8, 2 and 5 chances (out of 10) respectively of the 50-50 die-roll under the four possible compositions of Deck 3. As the second and third possible compositions of that deck are equally likely, ambiguity aversion requires preference for g_1 over g_2 and ambiguity seeking the reverse preference. We use this fact to classify subjects by ambiguity attitude. Subjects who were indifferent between g_1 and g_2 were classified as ambiguity neutral; and all of the remainder as ambiguity sensitive, with the latter group divided into ambiguity seeking and ambiguity averse. Since the predictions in relation to preference over g_1 and g_2 are common to the smooth ambiguity and α-MEU models, this procedure is in line with and neutral between both two models.

A potential qualification to this procedure is that, strictly, preference over g_1 and g_2 only determines a subject's attitude to ambiguity when the subject weights the second and third possible compositions of Deck 3 equally. We review robustness of our findings to violation of this condition and to other variations on our classification procedure in Appendix A.

3.8 Predictions and control

We now put the theoretical predictions in the context of the design. For the 2-ball condition, which matches the set-up of Section 2.2, the smooth ambiguity model predicts that those subjects who prefer g_1 to g_2 (ambiguity averse) should also prefer *mix* to each of f_1 and f_2; those who prefer g_2 to g_1

(ambiguity-seekers) should also prefer each of f_1 and f_2 to *mix*; and those indifferent between g_1 and g_2 (ambiguity neutral) should be indifferent between *mix*, f_1 and f_2. In contrast, the α-MEU model predicts that all subjects should be indifferent in the 2-ball condition between *mix* and each of its constituents, regardless of their preference over g_1 and g_2.

We use the 1-ball condition as a control in several related ways. In the 1-ball condition, the smooth ambiguity model joins the α-MEU model in predicting indifference between f_1, f_2 and *mix* as, in each possible composition of Deck 2, the number of spades equals the number of hearts, making the overall chances of receiving €20 the same under those three acts. If we observe preference for *mix* over its constituents among ambiguity averse subjects in the 2-ball condition, and if the smooth ambiguity model correctly diagnoses the only source of that preference, the preference should be absent in the 1-ball condition. However, it is possible that subjects will be attracted (or repelled) by the mixed act relative to its constituent acts for reasons other than the hedging argument of the smooth ambiguity model. For example, subjects might have an attitude, positive or negative, towards the presence in the resolution of the mixed act of another source of uncertainty, die-rolling, in addition to the drawing of cards from decks. But, if so, this should show up in both the 2-ball and the 1-ball conditions. Thus, the difference between the two conditions is of particular interest, regardless of whether we observe the predicted indifference in the 1-ball condition.

To build on these points, we now define variables used in our data analysis. We use $CE(f, C)$ to denote the certainty equivalent of act f in condition C (though we omit the condition where obvious from the context) and we use $AvCE(f, g, C)$ to denote the (arithmetic) mean of a subject's certainty equivalents for acts f and g in condition C. The following *premium variables* can then be defined for each subject:

Mixed act premium (2-ball)	=	$CE(mix,$ 2-ball$) - AvCE(f_1, f_2,$ 2-ball$)$;
Mixed act premium (1-ball)	=	$CE(mix,$ 1-ball$) - AvCE(f_1, f_2,$ 1-ball$)$;
2-ball premium	=	$CE(mix,$ 2-ball$) - CE(mix,$ 1-ball$)$;
Difference between mixed act premia	=	*Mixed act premium (2-ball)– Mixed act premium (1-ball).*

Mixed act premium (2-ball) measures the excess attractiveness of *mix* over its constituents in the condition where the smooth ambiguity model makes its distinctive prediction that ambiguity averse subjects prefer the mixed act and ambiguity seekers the constituent acts.[21] Mixed act premium (1-ball) measures the corresponding excess attractiveness in the condition where both models predict that all three types are indifferent between *mix* and its constituents. The variable "difference between mixed act

[21] Some readers may think it arbitrary to use $AvCE(f_1, f_2,$ 2-ball$)$ in cases where $CE(f_1,$ 2-ball$) \neq CE(f_2,$ 2-ball$)$. However, cases where $CE(f_1,$ 2-ball$) \neq CE(f_2,$ 2-ball$)$ are quite rare and we discuss a theoretically-grounded way of dealing with them in Appendix A. It turns out to make little difference to the results.

premia" measures how far "excess attractiveness" of *mix* over its constituents is greater in the 2-ball condition than it is in the 1-ball condition. Thus, it measures the influence of the hedging of independent ambiguities consideration, controlling for any other factors that (contrary to both models being considered) may make *mix* either more or less attractive than its constituent acts in the 1-ball condition. Finally, the 2-ball premium measures directly the extent to which *mix* is more attractive when it does offer a hedge across independent ambiguities than when it does not.

According to the smooth ambiguity model, all of these premium variables should be positive for the ambiguity averse, zero for the ambiguity neutral, and negative for the ambiguity seeking, except for mixed act premium (1-ball) which should be zero for all three types. The predictions of the α-MEU model are simply that each of the four premium variables should be zero for all types. Finally, SEU theory implies ambiguity neutrality and zero values of all four premium variables.[22] Thus, three of the premium variables discriminate between models (for ambiguity sensitive subjects): mixed act premium (2-ball); 2-ball premium; and difference between mixed act premia. The first is a direct comparison of the mixed act and its constituents in the 2-ball condition; the other two make comparisons across conditions, so exploiting the 1-ball control. The difference between mixed act premia is our most refined discriminator between the smooth ambiguity and α-MEU models: it measures the contribution of the hedging of separately resolving ambiguities motive to preference over *mix* and its constituents, while controlling for other motives that might also affect that preference.

4. Results

4.1 Preliminaries

The experiment was conducted at the University of Tilburg. 97 subjects took part, all of whom were students of the university.[23] They were paid a show-up fee of €5 on top of their earnings from the tasks, yielding a total average payment of €15.74. The main function of the risky acts resolved with Deck 1

[22] Except where otherwise stated, we consider subjects categorized as ambiguity averse, ambiguity seeking and ambiguity neutral separately when testing predictions for a given premium variable statistically. This takes account of the fact that, when the premium is not predicted to be zero, it is predicted to differ from zero in a *specified* (category-dependent) *direction* for each ambiguity-sensitive category. In these cases, we test the null hypothesis that the (central tendency of the) premium variable does not differ from zero (in the predicted direction) against the alternative that it does differ from zero in the direction predicted by the smooth ambiguity model, using a one-tailed test. In cases where all the theories we consider predict the premium will be zero, we test a null of no difference from zero against the alternative that there is such a difference in either direction, using a two-tailed test. For related tests of relationships between the underlying certainty-equivalents, we adopt the analogous policy: where the smooth ambiguity model predicts a difference between two certainty equivalents in a specified direction, for a given category of subject, we use a one-tailed test; otherwise we use a two-tailed test. Though theory mandates one-tailed tests in the cases where we use them, almost all our statistically significant results in one-tailed tests would also be statistically significant at the 5% level on a two-tailed test.

[23] We exclude from these figures and our data analysis five subjects who always chose the same option in at least one choice-list, so revealing misunderstanding by violating dominance.

was to enhance subjects' understanding of subsequent ones, but we report that the median certainty equivalents for 70% and 30% chances, respectively, of €20 were €11.73 and €5.58, suggesting levels of risk aversion not uncommon among experimental subjects. We now turn to ambiguous acts.

4.2 Results on classification of subjects

Certainty equivalents for g_1 and g_2 allow us to categorize subjects into three *types*: the ambiguity averse ($CE(g_1) > CE(g_2)$); the ambiguity neutral ($CE(g_1) = CE(g_2)$); and the ambiguity seeking ($CE(g_1) < CE(g_2)$). Out of a total of 97, the numbers of subjects of each type were 31, 50 and 16 respectively.

Although some studies have found a higher proportion of ambiguity sensitive subjects than we do, Ahn *et al*. (2014, p. 206) found that 72.7% of their subjects were either ambiguity neutral or close to it and Charness *et al.* (2013, p.3) found 60.3% of theirs to be ambiguity neutral. Thus, our findings are not out of line with the range of previous findings. Recall that our design was constructed to make it "easy" to reveal indifference between certain sets of acts the certainty equivalents of which were elicited side-by-side on the same screen. A subject who saw a relationship between two such acts which she regarded as making them equally attractive would have had no difficulty in giving certainty equivalents that reflected that judgment. From this perspective, the proportion of subjects coded as ambiguity neutral is actually quite encouraging, even though it lowers the proportion coded as ambiguity sensitive. As subjects clearly were able to express indifference between g_1 and g_2, there is no reason to think they would not have been able to do so between *mix* and its constituents (the certainty equivalents of which were also elicited side by side on the same screen) if they saw fit.[24]

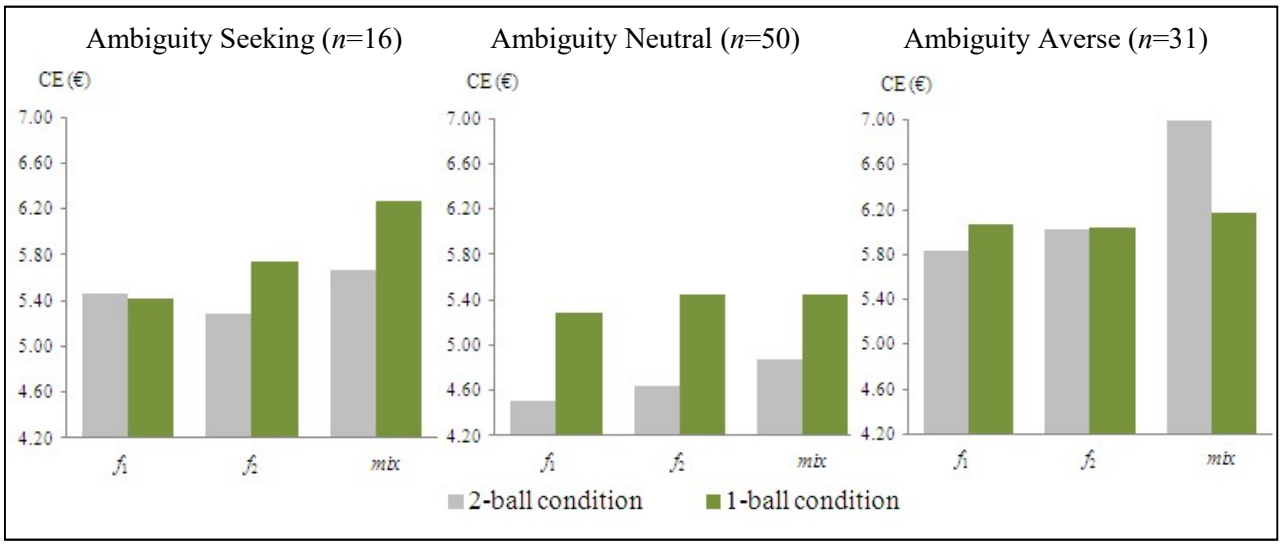
Notwithstanding ambiguity neutral being the largest group, the mean difference $CE(g_1) - CE(g_2)$ was €0.45 across all subjects, reflecting some ambiguity aversion on average. The corresponding figures for the two ambiguity sensitive types were €1.90 for the ambiguity- averse and –€0.93 for the ambiguity seeking.

4.3 Comparing certainty equivalents: central tendencies

As an initial display of our findings, Figure 3 reports the mean certainty equivalents, for each ambiguous act under each condition, separately by type of subject.

[24] In further support of this: Out of 97 subjects, 71 gave *identical* certainty equivalents for f_1 and f_2 in the 2-ball condition, in line with theoretical predictions. (The others are considered in Appendix A1.) For comparison, the only case we have where all theories would predict a difference between certainty equivalents is that of the two risky acts; here, just 7 subjects gave identical certainty equivalents.

Figure 3: Mean CE's for Ambiguity Seeking, Neutral, and Averse Subjects



As explained in Section 3.8, the most important features of our data are the premium variables defined in terms of the certainty equivalents. The mean, median and standard deviations of each of the four premium variables are reported Table 5.

Table 5: Premia (in €, rounded to nearest cent)

	Ambiguity Seeking (*n*=16)			Ambiguity Neutral (*n*=50)			Ambiguity Averse (*n*=31)		
	Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.
Premia									
Mixed act (2-ball)	0.30	0.46	2.68	0.29	0.00	1.58	1.06	0.73	2.63
Mixed act (1-ball)	0.69	0.00	2.53	0.09	0.00	2.02	0.12	0.00	2.39
2-ball	-0.60	-0.43	1.85	-0.58	0.00	2.13	0.83	0.55	2.53
Difference between mixed act premia	-0.39	-0.64[25]	2.01	0.21	0.00	2.00	0.95	0.30	2.39

Several points stand out from Figure 3 and Table 5. If, first, we confine attention to subjects coded as ambiguity averse, then the findings are, at eyeball level, very much in line with the predictions of the smooth ambiguity model. In particular, the right hand panel of Figure 3 shows that, for these subjects,

[25] Among the ambiguity seeking, median difference between mixed act premia is negative, though median mixed act premium (2-ball) is positive and median mixed act premium (1-ball) zero. This is not a typo, but just a sharper reflection of the fact, also evident for the ambiguity averse, that median (*x*-*y*) may not equal (median *x*) – (median *y*).

mix seems to have been judged on average notably more attractive than its constituents in the 2-ball condition, but not in the 1-ball condition. Table 5 indicates that, for the ambiguity averse, the mixed act premium (2-ball) and the 2-ball premium are both, on average and by median, positive and seemingly non-trivial; whereas the central tendencies of mixed act premium (1-ball) are close to zero.[26]

For ambiguity averse subjects, Wilcoxon signed-rank tests reveal that $CE(mix,$2-ball) exceeds each of $CE(f_1,$2-ball) and $CE(f_2,$2-ball) ($p = 0.007$ and $p = 0.008$, respectively) and also that $CE(mix,$2-ball) is larger than $CE(mix,$1-ball) ($p = 0.019$). In contrast, we cannot reject equality of $CE(mix$, 1-ball) with either $CE(f_1,$1-ball) or $CE(f_2,$1-ball) ($p = 0.664$ and $p = 0.635$, respectively). Thus, there is evidence, at the level of central tendencies, in favor of the hypothesis that ambiguity averse subjects value the hedge against independent ambiguities that *mix* offers over its constituents in the 2-ball condition but that, as also predicted by the smooth ambiguity model, this attraction to *mix* disappears in the 1-ball condition, where the ambiguities are not independent.

In the case of subjects coded as ambiguity neutral, all theories agree. The medians of each of the premium variables are exactly as predicted by the theories. However, surprisingly, ambiguity neutral subjects seem from Figure 3 to prefer each of the acts in the 1-ball condition over the same act in the 2-ball condition, as Wilcoxon signed-rank tests confirm.[27] The reason for this is unclear but, one possibility is that some subjects are averse to greater numbers of possible compositions of the deck. Whatever the reason, as the effect favors the 1-ball version, it does not seem to indicate any factor that would contribute to our earlier finding that ambiguity averse subjects prefer the 2-ball version of *mix* over its constituents. Indeed, if anything, it strengthens that finding.

Our findings for subjects coded as ambiguity seeking are more mixed than those for the ambiguity averse. For example, for these subjects, the mean and median values of the mixed act premium (2-ball) both have the wrong sign from the perspective of the smooth ambiguity model. However, the picture changes if we consider the premium variables that use the 1-ball control. The means and medians of the 2-ball premium and the difference between mixed act premia all take the sign predicted by the smooth model. That said, these effects receive only very limited corroboration in statistical tests[28] so we cannot reject the predictions of the α-MEU model for ambiguity seeking subjects. Given the small number of such subjects, it would inevitably be difficult to detect any statistically reliable pattern in their behavior.

---

[26] In interpreting the median values in Table 5, recall that our design makes it easy to reveal certain indifferences and that we only have interval estimates of certainty equivalents, so median values of precisely zero are less surprising than they might otherwise seem.

[27] p-values for rejection of null hypotheses $CE(mix,$1-ball) = $CE(mix,$2-ball), $CE(f_1,$2-ball) = $CE(f_1,$1-ball) and $CE(f_2,$2-ball) = $CE(f_2,$1-ball) are, respectively, 0.029, 0.010, and 0.001.

[28] Statistically, $CE(mix,$2-ball) is not significantly smaller than either $CE(f_1$, 2-ball) or $CE(f_2$, 2-ball) ($p = 0.226$ and $p = 0.121$, respectively)). The difference between $CE(mix,$2-ball) and $CE(mix,$1-ball) is only marginally significant even on the appropriate one-sided test ($p = 0.085$); and difference between mixed act premia is not ($p = 0.250$).

However, ambiguity averse and ambiguity seeking categories can be pooled, using transformations of the three premium variables that discriminate between models in predictions for ambiguity-sensitive subjects. For each of these variables, the transformation makes deviations from zero in the direction predicted by the smooth ambiguity model positive, and deviations in the opposite direction negative, by multiplying the original premium variable by minus 1 for the ambiguity seeking (only). Then, the smooth model predicts a positive value of the transformed variable for any ambiguity sensitive subject, whereas the α-MEU model predicts a zero value, and a negative value is possible but not predicted by either model. This transformation allows statistical tests to be conducted on the n = 47 ambiguity sensitive subjects taken as a single group. We find that the transformed mixed act premium (2-ball) is only marginally significantly larger than zero (p = 0.061), but the transformed 2-ball premium and transformed difference in mixed act premia are both significantly larger than zero (p = 0.013 and p = 0.013, respectively), so that, again, exploiting the 1-ball control sharpens the picture.

4.4 Categorical analysis

The analysis of the previous subsection is subject to two limitations. Firstly, it concentrates on *magnitudes* of certainty equivalents and premium variables, whereas the theoretical predictions are really about *ordinal* comparisons of certainty equivalents (and hence only about signs of the premium variables). Secondly, as it focuses on the "typical" subject in each type, it does not fully capture the proportion of subjects in a given type conforming to a given prediction. In this sub-section, we present a brief categorical analysis that addresses these points.

Table 6: "Signs" of Premia and Ambiguity Attitude

		$CE(g_1) - CE(g_2)$ Ambiguity attitude		
		< 0 (Seeking)	0 (Neutral)	> 0 (Averse)
Mixed act premium (2-ball)	>0	9	13	21
	0	1	26	3
	<0	6	11	7
2-ball premium	>0	4 (4)	12 (7)	19 (17)
	0	0 (4)	15 (23)	2 (6)
	<0	12 (8)	23 (20)	10 (8)
Difference between mixed act premia	>0	5 (5)	16 (16)	19 (17)
	0	1 (1)	19 (21)	4 (7)
	<0	10 (10)	15 (13)	8 (7)

We define (with slight abuse of terminology) the *sign* of a variable as taking one of three values: strictly positive, zero, or strictly negative. Table 6 presents contingency tables for the sign of $CE(g_1) - CE(g_2)$ (i.e. the subject's type) against the sign of each of the three premium variables that discriminate

between models. The table gives the frequencies of subjects with each combination of type and sign of premium variable, for each premium variable. The first number in each cell gives the absolute number, for each frequency.

As it may be "difficult" for a subject to achieve a value of precisely zero for a given premium variable, we also consider an alternative coding. We have already argued that subjects seemed to have no difficulty in achieving $CE(g_1) = CE(g_2)$, as these two certainty equivalents were elicited side-by-side on the same screen. For this reason, we use a requirement of exact equality here when classifying subjects as ambiguity neutral. But, that argument is less compelling for some of the premia. To achieve either a 2-ball premium of zero or a difference between mixed act premia of zero requires suitable alignment of certainty equivalents elicited across different screens. In view of these points, Table 6 also indicates parenthetically, for these variables, the frequencies under a revised coding scheme in which a sign of zero is attributed to the premium variable if its absolute value is no more than €0.20 from zero (an allowance equivalent to 4 rows of a zoomed-in choice list). Unsurprisingly, this pulls more observations into the central rows of the relevant panels of Table 6.

According to the smooth ambiguity model, each subject's type should match the sign of their premium variable, for each of the three premium variables presented in Table 6. To capture the extent of conformity with this prediction, we calculate, for each of these premium variables, the *sign-matching rate*, defined as the percentage of subjects for whom type matches the sign of the premium variable. Correspondingly, for each premium variable, we also calculate the *sign-zero rate*, defined as the percentage of subjects for whom the sign of the premium variable is coded as 0, in accordance with the α-MEU model.

Table 7 reports both rates, for each of the premium variables from Table 6, separately for all subjects; ambiguity sensitive subjects; and ambiguity averse subjects. Rates are given to the nearest percentage. As with Table 6, un-parenthesized entries correspond to the stricter coding rule for a zero sign on the premium variable and parenthesized entries to the looser coding. By construction, the looser coding rule for a sign of zero on the premium variable cannot lower the sign-zero rate. In fact, as Table 7 shows, it raises that rate in all cases and in some cases substantially so. In contrast, the looser coding rule sometimes raises and sometimes lowers the sign-matching rate; and most of these adjustments are quite small. In terms of the comparative performance of the smooth ambiguity and α-MEU models for a given premium variable and group of subjects, what matters is the difference between the sign-matching and the sign-zero rate. This is reduced by the looser coding rule for the latter rate in all cases shown.

Table 7: Sign-matching and Sign-zero Rates (%) by Premium Variable

Premium variable	Rate: Sign-…	All (n=97)	Ambiguity sensitive (n=47)	Ambiguity averse (n=31)
Mixed act (2-ball)	Matching	55	57	68
	Zero	31	9	10
2-ball	Matching	47 (49)	66 (53)	61 (55)
	Zero	18 (34)	4 (21)	6 (19)
Diff. between mixed act premia	Matching	49 (49)	62 (57)	61 (55)
	Zero	25 (30)	11 (17)	13 (23)

Using the looser of our codings where applicable, the sign-matching rate exceeds the sign-zero rate in every case reported in Table 7, by a margin never lower than 15 percentage points. If attention is restricted to ambiguity sensitive subjects then, using the coding that favors the sign-zero rate in the second and third cases, the sign-matching rate exceeds it by 48 (= 57 - 9) percentage points, 32 (= 53 - 21) percentage points and 40 (= 57 – 17) percentage points for the mixed-act premium (2-ball), the 2-ball premium, and the difference between mixed act premia, respectively. In this respect, the smooth ambiguity model outperforms the α-MEU model.

That said, the performance of the smooth ambiguity model in Table 7 is far from perfect. The sign-matching rates reported in the "All" column are only around 50%, and none of those reported in other columns exceeds 68%. (These figures compare with a benchmark of 33%, if the three values of sign were allocated at random.)

4.5: Main experiment: Summary of findings

The analysis of central tendencies reported in Section 4.3 and the individual-level categorical analysis of Section 4.4 broadly cohere with one another.

Where the smooth ambiguity and α-MEU models agree in relation to our design – that is, in their predictions for the ambiguity neutral in the 2-ball condition and for all types in the 1-ball condition – their shared predictions perform well, as judged by central tendencies, but less well at the individual-level. For example, the median value of each of our premium variables is zero in every case where both models predict that it will be zero, but neither model accounts for the fairly frequent incidence (evidenced by the central data column of Table 6) of individual-level violations of the shared prediction that each premium variable will be zero for each ambiguity neutral subject.

Our potential for discriminating between models is provided by ambiguity sensitive subjects, be they ambiguity averse or ambiguity seeking, since the two models disagree in their predictions for these subjects. Too few subjects are coded ambiguity seeking for a statistically significant pattern to emerge when that group is considered on its own but, when all ambiguity-sensitive subjects are pooled, the

three premium variables that distinguish between models tell broadly in favour of the smooth ambiguity model. This generalisation holds for individual-level analysis and central tendencies alike, but the evidence for it is less clear for the premium variable that only draws information from the 2-ball condition than it is for the two premium variables that exploit our 1-ball control by drawing information from both conditions. This qualification suggests that neither model fully captures the behaviour of the whole set of ambiguity-sensitive subjects in either condition taken separately, yet comparison of the 2-ball and 1-ball conditions is still broadly consistent with responses to separately resolving ambiguities in directions predicted by the smooth ambiguity model.

The qualification about reliance on measures that exploit the 1-ball control is not needed when attention is restricted to subjects coded as ambiguity averse as, for these subjects, all of the three premium variables that discriminate between models tell essentially the same story in both forms of analysis. We find clear and statistically significant patterns in the behaviour of the subjects coded as ambiguity averse that conform more closely to the predictions of the smooth ambiguity model than to those of the α-MEU model.

5. Extensions

Before concluding, we comment on some extensions of our investigations reported in this section and, especially, in Appendices A-C.

5.1 Robustness to categorization

As noted in Section 3.7, our categorization of subjects by ambiguity attitude assumes that they see the second and third compositions of Deck 3 as equally likely, as implied by the information provided. It also assumes that, even when they have this belief, they reveal ambiguity neutrality through exact *equality* between $CE(g_1)$ and $CE(g_2)$. Appendix A explores the robustness of our conclusions to relaxation of these assumptions: Appendix A.1 considers the possibility of subjects not realizing that the second and third compositions of Deck 3 are equally likely and Appendix A2 what happens if we allow closeness (rather than only equality) between $CE(g_1)$ and $CE(g_2)$ to count as indicating ambiguity neutrality. In each case, the details of our findings are affected but the general tenor is not.

A different possible concern about our analysis, in view of the findings of previous experiments in the literature, is its reliance on expected utility theory as the model of choice under objective risk. In this respect, our analysis is true to the smooth and α-MEU models, as usually formulated.[29] Notwithstanding this, in Appendix B, we show that our theoretical analysis would be robust to a

[29] Dean and Ortoleva (2017) introduces a model with close affinity to the MEU model that also captures violations of expected utility theory under objective risk, such as the Allais paradoxes, and explores the links between hedging, Ellsberg-like, and Allais-like behavior.

reformulation of these models in which expected utility theory is replaced in this role by any one of a range of non-expected utility models.

5.2 Divide across models: theory and a follow-up experiment

In Appendix C, we substantiate Section 1's assertion that the α-MEU and smooth ambiguity models stand as examples of broader classes of models of ambiguity sensitive preference: that is, respectively, of models of preferences whose representations focus exclusively on minimum and/or maximum expected utilities and of models whose representations also give weight to non-extremal expected utilities. In particular, we identify other models that, under important but plausible assumptions, share the predictions of models that we have considered in the main text in the context of our design.

In the case of the main study reported above, one model – that of Seo (2009) – shares the predictions of KMM's smooth ambiguity model in a particularly direct way. But, as footnote 7 explains, the two models differ in subtle respects, in particular in tying or not tying ambiguity aversion to violation of ROCL. This issue connects to an important debate in the ambiguity literature about whether sensitivity towards ambiguity is simply the result of non-standard reduction of compound lotteries. The experimental data of Halevy (2007) suggest that this is the case[30]: ambiguity aversion is strongly associated with violations of ROCL. Abdellaoui *et al.* (2015), however, observe that ambiguity-neutral subjects also often violate ROCL and that a significant proportion of subjects that do reduce compound lotteries are sensitive towards ambiguity, suggesting that ambiguity attitude is more than DMs violating ROCL.

In view of these points, we ran an additional experiment with two-stage objective lotteries in place of the ambiguous acts of the main study. In this follow-up study, the objective probabilities of the possible compositions of Decks 2 and 3 respectively were given to subjects. Specifically, the opaque bag that determined the composition of Deck 2 and Deck 3 was publicly filled with an equal number of balls numbered 1 and 4, so that subjects knew that each possible composition of Deck 2 was equally likely, as was each possible composition of Deck 3. In all other respects, the new treatment was identical to the main experiment.[31] Since the probabilities for each possible composition are objectively given, in the new experiment the behavior implied by MEU and KMM's model are identical and consistent with that implied by expected utility. As noted in footnote 7, in Seo (2009)'s model, violation of ROCL is implied by a concave ϕ and unaffected by whether the second-order uncertainty is an objective risk or not.

In total, 86 undergraduate students (average payment €15.24) from Tilburg University participated.[32] Classifying subjects based on the difference between the certainty equivalent of act g_1

[30] See also Chew *et al*. (2017).
[31] We used the tabular protocol. Instructions for the follow-up experiment are in Appendix E3.
[32] The data of 4 subjects who always chose the same option in at least one choice-list was removed from the analysis.

and g_2 yields the same general pattern of preferences as in the main experiment. In particular, 20 subjects reported $CE(g_1) > CE(g_2)$, 16 reported $CE(g_1) < CE(g_2)$, and 50 reported $CE(g_1) = CE(g_2)$. As all risks are objective in the follow-up experiment, the two strict inequalities are violations of ROCL and, in that sense, contrary to KMM's model and MEU. A two-sided Wilcoxon rank-sum test does not reveal a statistically significant difference between the two experiments in the statistic $CE(g_1) - CE(g_2)$ ($p = 0.231$). These findings are in line with the Seo (2009) model and also suggestive of the findings of Halevy (2007).

Just as with the main experiment, it is particularly interesting to consider the preferences over *mix* and its constituents of subjects appearing to be "ambiguity averse", in the sense that (($CE(g_1) > CE(g_2)$)).[33] But, here the picture is mixed, in part depending on whether the 1-ball control is used. For subjects with $CE(g_1) > CE(g_2)$ in the follow-up experiment, one-tailed Wilcoxon signed-rank tests provide some evidence that *mix* is preferred to its constituents, as $CE(mix,$2-ball) exceeds each of $CE(f_1,$2-ball) and $CE(f_2,$2-ball) ($p = 0.063$ and $p = 0.039$, respectively). However, this evidence disappears when measures exploiting the 1-ball control are used. In particular, the 2-ball premium is not significantly positive ($p = 0.340$) and nor is the difference between mixed act premia ($p = 0.463$). Recall in contrast that, in the main experiment, all premium variables told the same story for the ambiguity averse. This difference between the findings of the two experiments is more supportive of KMM's version of the smooth model than of that of Seo (2009).

Moreover, in the follow-up experiment, a Friedman test does not detect differences between the distributions of the ranks of the certainty equivalents of the acts f_1, f_2 and *mix* in the 1-ball condition, f_1, f_2 and *mix* in the 2-ball condition and g_1 and g_2 in the 2-ball condition ($p = 0.657$). In contrast, a similar Friedman test for the main experiment reveals that the ranks of those certainty equivalents are different ($p = 0.001$). These findings suggest that the aggregate pattern of preference over these acts that we observe in the main experiment is not entirely driven by violations of ROCL in preferences over objective lotteries. However, two qualifications are in order. First, we do observe many violations of ROCL at the individual level in the follow-up experiment, so its findings do not in themselves support ROCL for objective risks. Secondly, as the follow-up study was run later than the main one, subjects were not randomly assigned between the two; so, any comparison of the studies must be made with caution. Partly for this reason, we focus on the main study in our final concluding comments.

6. Concluding Remarks

We have presented an experiment designed to discriminate empirically between two well-known and widely applied models of ambiguity attitude – the α-MEU and smooth ambiguity models – which stand

[33] It is necessary, in Seo's model, for ϕ to be concave for $CE(g_1) > CE(g_2)$. Therefore, in Seo's model, subjects with $CE(g_1) > CE(g_2)$ will prefer *mix* to its constituent acts in the 2-ball condition, but not in the 1-ball condition.

as representatives of a broader theoretical divide. This is important as the conclusions of existing applied economic models that incorporate ambiguity are often derived from specific models of ambiguity sensitivity, sometimes on one side of the divide we mark and sometimes on the other.

Our strategy for distinguishing between models focuses on whether or not an agent's preference for hedging separate ambiguities depends on ambiguity attitude. Interpreting our findings narrowly in terms of a comparison of the smooth ambiguity and α-MEU models, we find that, while neither model captures all aspects of our data, there are striking features of the data that conform more closely to the smooth ambiguity model than to the α-MEU model. As explained in Section 1 and Appendix C, this discrimination extends to a broader divide between models which do or do not give weight to non-extreme expected utilities, among those considered possible.

Finally, our results provide a more general reassurance to the theories of ambiguity aversion. It has long been argued, at a theoretical level, that reliance on stochastic mixing as a way to hedge ambiguity is a defining part of a rational response to ambiguity. Indeed, such is the motivating basis of Schmeidler's (1989) *Uncertainty Aversion* axiom, which lies at the heart of ambiguity averse preference models, quite generally (Cerreia-Vioglio *et al*., 2011). However, the best-known prior study that we are aware of which investigates the link between ambiguity aversion and preference for randomization – Dominiak and Schnedler (2011) – reports little support for this central premise of theories of ambiguity aversion. In view of this, it is particularly notable that our main finding, specifically strict preference for *mix* over f_1 and f_2 on the part of the ambiguity averse in the 2-ball condition, supports the Uncertainty Aversion axiom.[34] Hence, a further overall "take-away" from our findings, is that they provide evidence for a link between ambiguity aversion and propensity to use randomization to hedge ambiguity, and thus for the central foundational principle of theories of ambiguity aversion, quite generally.

[34] Consider ambiguity averse DMs. An α-MEU preference and a smooth ambiguity preference both satisfy Schmeidler's Uncertainty Aversion axiom. The axiom requires that there will be instances when, given two acts between which the DM is indifferent, the mixture between the acts is strictly preferred to either act, delineating the characteristic departure of ambiguity aversion from behavior satisfying the (Anscombe and Aumann) SEU theory. However, the axiom does not further stipulate what those instances will be. Our acts *mix*, f_1 and f_2 in the 2-ball condition provide an instance where a smooth ambiguity model implies a strict preference while an α-MEU model predicts indifference. (In contrast, both models predict strict preference for the mixture of two bets, one on each color of an Ellsberg 2-color ambiguous urn, over its constituents.) Hence, had we found that subjects coded as ambiguity averse were largely indifferent between our three acts in the 2-ball condition then that would be evidence neither in support nor against Schmeidler's axiom, whereas the strict preference that what we do find is positive support for the axiom. In our design, in line with the axiom, the objective randomization that *mix* offers is conducted as a final step in determining the outcome of the act. In contrast, in Dominiak and Schnedler's design, the objective randomization is ex ante.

References

Abdellaoui, M., Klibanoff, P. and Placido, L. (2015) "Experiments on compound risk in relation to simple risk and to ambiguity," *Management Science*, 61(6), 1306-1322.

Ahn, D. S. (2008) "Ambiguity without a state space," *Review of Economic Studies*, 75, 3-28.

Ahn, D.S., S. Choi, D. Gale and S. Kariv (2014) "Estimating Ambiguity Aversion in a Portfolio Choice Experiment," *Quantitative Economics*, 5, 195-223.

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

Andreoni, J., Schmidt, T. and Sprenger, C. (2014) "Measuring ambiguity aversion: experimental tests of subjective expected utility," paper available at http://www.eea-esem.com/files/papers/EEA-ESEM/2014/1588/Measuring%20Ambiguity.pdf

Anscombe, F.J. and R.J. Aumann (1963) "A definition of subjective probability," *Annals of Mathematical Statistics,*" 34, 199–205.

Attanasi, G., C. Gollier, A. Montesano and N. Pace (2014) "Eliciting ambiguity aversion in unknown and in compound lotteries: A smooth ambiguity model experimental study," *Theory and Decision*, 77, 485-530.

Azrieli, Y., C.P. Chambers and P. Healy (2016) "Incentives in experiments: a theoretical analysis," Working paper, University of California, San Diego, forthcoming in *Journal of Political Economy*.

Bade, S. (2015) "Randomisation devices and the elicitation of ambiguity-averse preferences," *Journal of Economic Theory*, 159, 221-35.

Baillon, A. and Bleichrodt, H. (2015) "Testing ambiguity models through the measurement of probabilities for gains and losses," *American Economic Journal: Microeconomics*, 7, 77-100.

Baillon, A, Halevy, Y. and Li, C. (2015) "Experimental elicitation of ambiguity attitude using the random incentive system" paper available at http://faculty.arts.ubc.ca/yhalevy/RIS_Ambiguity.pdf.

Baltussen, G., G.T. Post, M.J. van den Assem and P.P. Wakker (2012) "Random incentive systems in a dynamic choice experiment," *Experimental Economics*, 15, 418-43.

Bardsley, N., Cubitt, R.P., Loomes, G., Moffatt, P., Starmer, C., and R. Sugden (2010) *Experimental Economics: Rethinking the Rules*, Princeton University Press.

Beattie, J. and Loomes, G. (1997) "The impact of incentives upon risky choice experiments", *Journal of Risk and Uncertainty*, 14, 149-62.

Becker G., DeGroot, M. and J. Marschak (1964) "Measuring utility by a single-response sequential method," *Behavioral Science*, 9, 226-232.

Bosch-Domenech, A. and J. Silvestre (2013) "Measuring risk aversion with lists: a new bias," *Theory and Decision*, 75, 465-96.

Bossaerts, P., S. Guarnaschelli, P. Ghirardato and W. Zame (2009) "Ambiguity and asset prices: an experimental perspective," *Review of Financial Studies*, 23, 1325–1359.

Camerer, C. and M. Weber (1992) "Recent developments in modeling preferences: uncertainty and ambiguity," *Journal of Risk and Uncertainty*, 5, 325-370.

Cason, T. and C. Plott (2014) "Misconceptions and game form recognition: challenges to theories of revealed preference and framing," *Journal of Political Economy,* 122, 1235-70.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M., and L. Montrucchio (2011) "Uncertainty averse preferences," *Journal of Economic Theory*, 146, 1275-1330.

Chew, S. H., Miao, B. and Zhong, S. (2017) "Partial Ambiguity," *Econometrica*, 85, 1239–1260.

Chambers, R. G., and T. Melkonyan (2017) "Ambiguity, reasoned determination, and climate-change policy," *Journal of Environmental Economics and Management*, 81, 74-92.

Charness, G, Karni, E. and D. Levin (2013) "Ambiguity attitudes and social interactions: an experimental investigation," *Journal of Risk and Uncertainty*, 46, 1-25.

Collard, F., Mukerji, S, Tallon, J.M. and K. Sheppard (2018) "Ambiguity and the historical equity premium," mimeo, forthcoming in *Quantitative Economics*.

Conte, A. and J.D. Hey (2013) "Assessing multiple prior models of behaviour under ambiguity," *Journal of Risk and Uncertainty*, 46, 113-132.

Cubitt, R.P., van de Kuilen, G. and S. Mukerji (2018) "The strength of sensitivity to ambiguity", *Theory and Decision*, forthcoming as hard copy, published 'Online First' at http://link.springer.com/article/10.1007/s11238-018-9657-9.

Cubitt, R.P., Starmer, C. and R. Sugden (1998) "On the validity of the random lottery incentive system," *Experimental Economics*, 1, 115-31.

Dean, M. and P. Ortoleva (2017) "Allais, Ellsberg, and preferences for hedging," *Theoretical Economics*, 12, 377-424.

Dimmock, S.G., Kouwenberg, R., Mitchell, O.S. and Peijnenburg, K. (2015) "Estimating ambiguity preferences and perceptions in multiple prior models: evidence from the field," *Journal of Risk and Uncertainty*, 51, 219-244.

Dominiak, A, and W. Schnedler (2011) "Attitudes toward uncertainty and randomization: an experimental study," *Economic Theory* 48, 289-312.

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

Epstein, L.G. (2010) "A paradox for the 'smooth ambiguity' model of preference," *Econometrica*, 78, 2085-2099.

Epstein, L.G. and M. Schneider (2010) "Ambiguity and asset markets," *Annual Review of Financial Economics*, 2(1), 315-346.

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

Etner, J., Jeleva, M. and Tallon, J.M. (2012) "Decision theory under ambiguity," *Journal of Economic Surveys*, 26, 234-70.

Fischbacher, U. (2007) "Z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10, 171–178.

Gajdos, T., T. Hayashi, J.-M. Tallon and J.-C. Vergnaud, (2008) "Attitude toward imprecise information," *Journal of Economic Theory*, 140, 23-56.

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

Gilboa, Itzhak, and Massimo Marinacci (2013) "Ambiguity and the Bayesian paradigm," in D. Acemoglu, M. Arellano and E. Dekel (eds.) *Advances in Economics and Econometrics, Tenth World Congress*, vol. 1, Cambridge University Press, Cambridge, United Kingdom.

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

Gul, F. and W. Pesendorfer (2014) "Expected uncertain utility theory," *Econometrica*, 82(1), 1-41.

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

Hayashi, T. and R. Wada (2009) "Choice with imprecise information: an experimental approach," *Theory and Decision*, 69, 335–373.

Hey, J.D. and J. Lee (2005) "Do subjects separate (or are they sophisticated)?", *Experimental Economics,* 8, 233-265.

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

Hey, J.D. and N. Pace (2014) "The explanatory and predictive power of non two-stage-probability theories of decision making under ambiguity," *Journal of Risk and Uncertainty*, 49, 1-29.

Holt, C.A. (1986) "Preference reversals and the independence axiom," *American Economic Review*, 76, 508-514.

Hurwicz, L. (1951) "Optimality criteria for decision making under ignorance," *Cowles Commission Papers*, 370.

Jaffray, J.Y. (1989) "Linear utility theory for belief functions," *Operations Research Letters,* 8, 107-112.

Jahan-Parvar, M.R., and H. Liu (2014) "Ambiguity aversion and asset prices in production economies," *The Review of Financial Studies*, 27(10), 3060-3097.

Ju, N., and J. Miao (2012) "Ambiguity, learning, and asset returns," *Econometrica*, 80(2), 559-591.

Kahneman, D. and A. Tversky (1979) "Prospect theory: an analysis of decision under risk," *Econometrica*, 47, 263-91.

Klibanoff, P., M. Marinacci and S. Mukerji (2005) "A smooth model of decision under uncertainty," *Econometrica*, 73, 1849–1892.

Klibanoff, P., M. Marinacci, and S. Mukerji (2012) "On the smooth ambiguity model: a reply," *Econometrica*, 80, 1303-23.

Kothiyal, A., Spinu V. and Wakker P. (2014) "An experimental test of prospect theory for predicting choice under ambiguity," *Journal of Risk and Uncertainty*, 48, 1-17.

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

Machina, M.J. and D. Schmeidler (1992) "A more robust definition of subjective probability," *Econometrica*, 60, 745–780.

Marinacci, M. (2015) "Model uncertainty," *Journal of the European Economic Association*, 13, 1022-1100.

Millner, A., Dietz, S., and G. Heal (2013) "Scientific ambiguity and climate policy," *Environmental and Resource Economics*, 1-26.

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

Neilson, W.S. (2010) "A simplified axiomatic approach to ambiguity aversion," *Journal of Risk and Uncertainty*, 41, 113–124.

Olszewski, W.B. (2007) "Preferences over sets of lotteries," *Review of Economic Studies,* 74, 567-595.

Quiggin, J. (1982) "A theory of anticipated utility," *Journal of Economic Behaviour and Organization*, 3, 323–343.

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

Seo, K. (2009) "Ambiguity and second-order belief," *Econometrica*, 77, 1575–1605.

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

Starmer, C. (2000) "Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk," *Journal of Economic Literature,* 38, 332-382.

Starmer, C. and Sugden, R. (1991) "Does the random-lottery system elicit true preferences?" *American Economic Review* 81: 971-978.

Thaler, R. and E. Johnson (1990) "Gambling with the house money and trying to break even: the effect of prior outcomes on risky choice," *Management Science*, 36, 643-660.

Trautmann, S.T. and van de Kuilen, G. (2016) "Ambiguity attitudes," in G. Keren and G. Wu (eds.) *The Wiley-Blackwell Handbook of Judgment and Decision Making*.

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

Wakker, P.P. (2010) *Prospect Theory for Risk and Ambiguity*, Cambridge University Press, Cambridge, United Kingdom.

Appendix A: Robustness to classification

In this appendix, we consider the robustness of our findings to variations in our procedure for classifying subjects by ambiguity attitude, as explained in Section 5.

A.1 Robustness towards beliefs

As explained in Section 3.7, our classification of subjects as ambiguity averse, ambiguity neutral or ambiguity seeking used a procedure that, strictly, relies on subjects being indifferent, in the 2-ball condition, between f_1 and f_2. In fact, for 71 subjects, the condition $CE(f_1,$2-ball) = $CE(f_2,$2-ball) holds *exactly*. For the remaining 26 subjects, the statistic $CE(g_1) - CE(g_2)$ is well-defined, but no longer sufficient to identify a subject's ambiguity attitude, as the subjects do not appear to weight the second and third possible compositions of Deck 3 equally. In this appendix we comment on how that affects our analysis, in particular, the robustness of our conclusions regarding the comparison of models.

For purposes of this discussion, we fix the domain of preferences to include those described by:

1. an α-MEU model with the set of probabilities in the representation given by $\mathcal{P} = \{p_i \mid i = 1,\ldots,4\}$, or its convex hull, where p_i refers to the probabilities described in Table 2;
2. an SEU model (specifically, an Anscombe-Aumann model) which puts positive weight on each of the four probabilities, p_i, $i = 1,\ldots,4$;
3. a smooth ambiguity model with a non-linear ϕ and a μ which puts positive weight on each of the four probabilities, p_i, $i = 1,\ldots,4$.

Within this domain, the α-MEU model predicts indifference between f_1 and f_2 and *mix* regardless of preference over g_1 and g_2. In contrast, both the smooth ambiguity and SEU models allow non-indifference between f_1 and f_2 if p_2 and p_3 are weighted unequally.[35] Importantly, they also require that the directions of a subject's preference over g_1 and g_2 and her preference over *mix* and f_2 match, as can seen from an inspection of the relevant rows of Table 3. However, SEU imposes an additional restriction that the smooth model does not, namely, that the direction of preference between f_1 and f_2 must also match that between g_1 and g_2. Under the smooth ambiguity model, non-linearity of ϕ could upset this correspondence, thereby distinguishing SEU from smooth ambiguity. In fact, of the 26 subjects who are not indifferent between f_1 and f_2 in the 2-ball condition, only 5 conform to the additional restriction imposed by SEU.

Thus, if we set SEU aside and focus on the two models of ambiguity sensitivity, we may use a comparison of the preferences between *mix* and f_2 and between g_1 and g_2 as a means of discriminating between the models, even when the subject is not indifferent between f_1 and f_2. Hence, our strategy for comparing the α-MEU and smooth ambiguity models is a simple modification of our analysis in the main part of the paper. We replace $AvCE(f_1, f_2,$ 2-ball) with $CE(f_2,$ 2-ball) in the definitions of mixed act premium (2-ball) and of difference between mixed act premia. For a given sign of $CE(g_1) - CE(g_2)$,

[35] Allowing this possibility involves relaxing our maintained hypothesis that subjects understand the implications of their information, but is consistent with continuing to assume that they believe what they are told directly.

the predictions of the smooth ambiguity model for the signs of the two re-defined premium variables are exactly as for the original definitions. Similarly, for the α-MEU model, the implications for the signs of the two re-defined premium variables are exactly as for the original definitions, since the model predicts indifference between *mix* and f_2 for any value of $CE(g_1) - CE(g_2)$. These points are unaffected by whether $CE(g_1) - CE(g_2)$ indicates ambiguity attitude.

The redefinition can only affect subjects for whom $CE(f_1, 2\text{-ball}) \neq CE(f_2, 2\text{-ball})$. (These subjects are spread roughly evenly across the three categories by sign of $CE(g_1) - CE(g_2)$.) The sign of the mixed act premium (2-ball) is changed by the redefinition in only nine cases and the sign of the difference between mixed act premia in only three. Clearly, the latter point makes little difference, so we focus on the former, using the stricter coding rule for a zero sign on a premium, because the redefinition of the mixed act premium (2-ball) makes it the difference of two certainty equivalents elicited next to each other. For the mixed act premium (2-ball), the number of subjects with a zero sign rises from 30 to 35, whereas the number with a sign matching that of $CE(g_1) - CE(g_2)$ falls from 53 to 51. Though this impact of the redefinition slightly favors the α-MEU model, the sign-matching rate for all subjects still exceeds the sign-zero rate by more than 16 percentage points. Thus, the main qualitative conclusions of the categorical analysis (Section 4.4) are not affected. Moving beyond signs, the median mixed act premium (2-ball) rises by 2 (resp. 0) euro cents among subjects with $CE(g_1) - CE(g_2) >$ (resp. =) 0. Hence, for these subjects the modification makes very little difference to the central tendency. But, among those with $CE(g_1) - CE(g_2) < 0$, the median mixed act premium (2-ball) rises from €0.46 to €0.85. This is a movement in the wrong direction for *both* smooth and α-MEU models. In that sense, it does not affect the conclusions that we drew from our analysis of the central tendencies of certainty equivalents in Section 4.3, about the *relative* performance of those models.

A.2 Robustness to trembles

In Section 4, we used exact equality of $CE(g_1)$ and $CE(g_2)$ to classify subjects based on their attitude towards ambiguity, because these certainty equivalents were presented to subjects side-by-side on the same screen. Here we consider whether our results are robust if we apply looser coding rules to classify subjects in terms of ambiguity attitude. In particular, we distinguish between three coding rules for monetary sums. Under strict coding (S), a variable of interest has to be exactly zero in order to be counted as zero; whereas, under Tremble-tolerant (TT) coding, the variable of interest has to be no more than + 0.05 and no less than – 0.05; and, under Loose (L), it has to be no more than + 0.20 and no less than – 0.20. In terms of choice behaviour, these codings translate respectively into no margin of error in a zoomed-in choice list (S), a margin of error equivalent to one row of a zoomed-in choice-list (TT), and a margin of error equivalent to four rows of a zoomed-in choice-list (L), respectively.

Table A1 and A2 replicate the results reported in Table 5, using Tremble-Tolerant and Loose coding respectively, when classifying subjects by attitude toward ambiguity, rather than the Strict

coding used in Table 5. By construction, applying a looser coding to type-classification reduces the number of ambiguity-sensitive subjects, specifically in our case from 47 under Strict coding (Table 5) to 39 under Tremble-tolerant coding (Table A1) and 36 under Loose coding (Table A2). We have already defended Strict coding for use in identifying ambiguity-neutrality, on the grounds that $CE(g_1)$ and $CE(g_2)$ are elicited right next to each other on the same screen. In view of this point, we also see Tremble-Tolerant as the more plausible of the two alternative coding rules for ambiguity attitude. But, we include both for completeness.

Table A1: Premia (in €, rounded to nearest cent) with Tremble Tolerant Coding

	Ambiguity Seeking (n=11)			Ambiguity Neutral (n=58)			Ambiguity Averse (n=28)		
	Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.
Premia									
Mixed act (2-ball)	0.89	1.05	2.69	0.24	0.00	1.66	1.03	0.69	2.77
Mixed act (1-ball)	0.97	0.00	2.98	0.10	0.00	1.95	0.07	0.13	2.43
2-ball	-0.20	-0.05	1.57	-0.59	-0.02	2.12	0.84	0.53	2.63
Difference between mixed act premia	-0.09	-0.70	1.65	0.13	0.00	2.06	0.95	0.29	2.46

Table A2: Premia (in €, rounded to nearest cent) with Loose Coding

	Ambiguity Seeking (n=11)			Ambiguity Neutral (n=61)			Ambiguity Averse (n=25)		
	Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.
Premia									
Mixed act (2-ball)	0.89	1.05	2.69	0.36	0.00	1.73	0.82	0.50	2.84
Mixed act (1-ball)	0.97	0.00	2.98	0.05	0.00	2.00	0.19	0.25	2.39
2-ball	-0.20	-0.05	1.57	-0.45	0.00	2.18	0.66	0.10	2.70
Difference between mixed act premia	-0.09	-0.70	1.65	0.31	0.00	2.19	0.63	0.15	2.33

Our main findings regarding ambiguity averse subjects replicate under the Tremble-Tolerant coding, despite the lower number of observations.[36] First, the central tendencies of the premium variables are hardly affected by switching from Strict to Tremble-Tolerant coding. Moreover, mixed act premium (2-ball) is again significantly positive ($p = 0.020$), whereas mixed act premium (1-ball) is not ($p = 0.319$). The difference between these premia is also significantly positive ($p = 0.049$). Although, under Loose coding, mixed act premium (2-ball) and mixed-act premium (1-ball) are lower than those under tighter codings and neither significantly different from zero, this may in part be due to low power and, in any case, we see the Tremble-Tolerant coding as more plausible for classification of subjects by type of ambiguity attitude. Nevertheless, the reader may be surprised that reducing the number of subjects coded as ambiguity-averse, by adopting Loose rather than Strict coding for ambiguity neutrality lowers the mean and median of the key premium variables for ambiguity averse subjects. It is important to note, however, that the predictions of the smooth ambiguity model for these variables are only ordinal. In particular, the model does not require that the larger is $CE(g_1) - CE(g_2)$ the larger will be the premium variables predicted to be positive for the ambiguity averse. Instead, the predictions are about signs.

In view of this, we turn to the robustness of our categorical analysis. For this, we need to consider combinations of coding rules, as we must code both type of ambiguity attitude and sign of the premium. So, for brevity, we focus – for each premium variable analysed in Sections 4.4 – on the robustness of our findings to different coding rule combinations of one measure of the comparative performance of the smooth ambiguity and α-MEU models, namely the excess of the sign-matching rate over the sign-zero rate for ambiguity sensitive subjects. Table A3 reports this statistic for each premium variable considered in Section 4.4, for four different coding combinations.[37] The label at the top of each column gives the coding combination that column reports, with the first word of the label giving the coding rule applied to ambiguity neutrality in the classification by type and the second word the coding rule applied in coding the sign of the premium. The third and fourth columns report new codings, not considered in the main text, whereas the first two columns repeat, for comparison, the two cases considered in Section 4.4 (where we insisted on Strict coding for type).

[36] We do not pursue statistical comparisons for ambiguity seeking subjects under the looser codings as there are only eleven such subjects.

[37] We consider Strict and Loose coding for signs of premia, as in the main text, and all three codings for type. We consider that it does not make sense to use a looser coding for type than for sign of premium, as type is obtained from certainty equivalents elicited next to one another.

Table A3: Sign-matching rate minus sign-zero rate (%) for ambiguity sensitive subjects: different codings

	Coding[a]			
	Strict-Strict (n = 47)	Strict-Loose (n = 47)	TT-Loose (n = 39)	Loose-Loose (n = 36)
Mixed act premium (2-ball)	49	38	31	25
2-ball premium	62	32	31	25
Difference between mixed act premia	51	40	36	31

[a]) Strict-Strict refers to Strict coding of types and signs of premia; Strict-Loose refers to Strict coding of types and Loose coding of signs of premia; TT-Loose refers to Tremble-Tolerant coding of types and Loose coding of signs of premia; Loose-Loose refers to Loose coding of types and signs of premia. Cell entries are rounded to the nearest integer.

As can be seen in Table A3, our main result that the smooth model outperforms the α-MEU model for ambiguity sensitive subjects is robust to the coding used, in the sense that in all cases shown the sign-matching rate exceeds the sign-zero rate by at least twenty-five percentage points. However, the result becomes less pronounced under looser coding as one moves from left to right along any row of the table, with both sample size and the excess of the sign-matching rate over the sign-zero rate falling.

Appendix B: Non-Expected Utility for Risk

A possible concern about our theoretical analysis of preference over the acts is its use of expected utility theory as the underlying model of choice under risk. In this respect, our analysis is true to the smooth ambiguity and α-MEU models as usually formulated. But, as there is evidence from many experiments that subjects deviate from expected utility theory under risk (Starmer, 2000), one might wonder how the theoretical analysis would be affected if this were true of our subjects too. It is possible to imagine more general formulations of these models which allow an induced probability distribution on consequences to be evaluated by a quite general non-expected utility functional (see, for example, the discussion of Corollary 1 in KMM). In this appendix, we show that our testing strategy for distinguishing between the α-MEU and the smooth ambiguity models, and our conclusions about the relative support of each model in the data, would be robust to such reformulation.

An act together with a *given* probability distribution on the state space induces a probability distribution on consequences. Since consequences may be lotteries, the induced distribution is, in general, a distribution over lotteries and hence, a two-stage lottery. Both the α-MEU and the smooth ambiguity model, in their standard formulations, evaluate such an induced distribution just as would an expected utility model (see, e.g., Klibanoff *et al.* (2012), Equation 1.1 and 1.2). In particular, the induced two-stage lottery is evaluated by reducing it to the corresponding one-stage lottery and computing the expected utility of the reduced lottery. We now consider a reformulation of the evaluation of such induced probability distributions that is more general than expected utility.

For brevity, we consider only a domain of induced probability distributions comprising one-stage and two-stage objective lotteries defined on the set of final monetary consequences used in our experiment, i.e. on {€20, €0}. Let preferences over such lotteries be represented by maximization of *any* real-valued function V, defined on the lotteries, such that (i) preferences respect first-order stochastic dominance; and (ii) V satisfies the following *Limited Reduction* condition: $V(\{\{€20, ½; €0, ½\}, r; \{€0,1\}, 1 - r\}) = V(\{€20, r/2; €0, 1 - (r/2)\}$ for any $1 > r > 0$. Notice that, Limited Reduction is much weaker than the standard reduction principle for compound lotteries and does not, *by itself*, impose any restriction on preferences over *one-stage* lotteries. It simply links preferences over two-stage lotteries of a particularly simple form to preferences over particular one-stage lotteries, in the way specified. This would be quite compatible, for example, with preferences over one-stage lotteries being generated by cumulative probability-weighting with a distorted (monotonic) weighting function.

We now generalize the α-MEU and smooth ambiguity models by using maximization of V, defined as above, in place of maximization of the expectation of u, as their representation of preferences with a *given* probability distribution on the state space. How are the acts considered in the experiment evaluated under this generalization? To compress notation, let $V(\{€20, 0.4; €0, 0.6\}) = x$, $V(\{€20, 0.25; €0, 0.75\} = y$, and $V(\{€20, 0.1; €0, 0.9\}\}) = z$. From first-order stochastic dominance, $x > y > z$.

Applying Limited Reduction where necessary, the values of V for each act and deck composition are given in Table B1 (to be compared with Table 3 in the main text).

Table B.1: Values of V

	p_1	p_2	p_3	p_4
f_1	z	z	x	x
f_2	z	x	z	x
mix	z	y	y	x
g_1	y	y	y	y
g_2	y	x	z	y

From here, the theoretical analysis of the 2-ball condition can proceed essentially just as in the main text. For the α-MEU model, as generalized in the preceding paragraph, the DM's preferences must satisfy $f_1 \sim f_2 \sim mix$. For the smooth ambiguity model, as generalized in the preceding paragraph (but imposing, as in the main text, that $\mu(p_2) = \mu(p_3)$), the DM's preferences must satisfy the conditions that $f_1 \sim f_2$; and that, for $f \in \{f_1, f_2\}$, $g_1 \succ g_2 \Leftrightarrow mix \succ f$; $g_1 \sim g_2 \Leftrightarrow mix \sim f$; and $g_1 \prec g_2 \Leftrightarrow mix \prec f$. Thus, the predictions of Section 2.2 about how preferences over acts in $\{f_1, f_2, mix\}$ are (or are not) related to preferences over $\{g_1, g_2\}$ under the two models are robust to the generalizations of the preceding paragraph. What matters for those predictions is just that preferences over one-stage lotteries satisfy first-order stochastic dominance and cohere with preferences over two-stage lotteries to the extent required by Limited Reduction. This is compatible with departures from expected utility theory in preferences over one-stage objective risk, even quite marked departures.

Appendix C: Divide Across Models

For concreteness in the main text, it was convenient to analyze and interpret the results purely in terms of the α-MEU model and the smooth ambiguity model. However, the divide between models addressed by our design is broader, as we explain in this appendix. Consider the formal representation of the setting in our experiment. Think of $\mathcal{P}$ as the set of probabilities (on the set of states $S = \{B_1, B_2, R_1, R_2\}$) in an α-MEU model, or the set of probabilities in the core of the convex capacity in the representation of a Schmeidler (1989) model of uncertainty aversion, or the support of μ in a smooth ambiguity model. Let Z denote a component, $Z = B, R$. With slight abuse of terminology let B_1 and R_1 be the "first" elements of their respective component, and B_2 and R_2 be the "second" elements. (For interpretation, B_1 equals spade, B_2 equals club, R_1 equals heart, R_2 equals diamond.) Let Z_i denote the *i*-th element from the Zth-component, $i = 1,2$. Denote the set of probabilities of drawing a first element from the Zth-component by $\Gamma_Z = \{p(Z_1): p \in \mathcal{P}\}$. Let $\mathcal{P}$ satisfy the following properties:

Property One: $\Gamma_R = \Gamma_B$.

Property Two: Γ_Z is non-singleton, for $Z = \{B, R\}$.

Property Three: If $q \in \Gamma_R$ and $q' \in \Gamma_B$, there is $p \in \mathcal{P}$ such that $p(R_1) = q$ and $p(B_1) = q'$.

Property Four: If $p \in \mathcal{P}$ then $p(\{Z_1, Z_2\}) = 0.5$.

In our experiment, in the 2-ball condition that implements the theoretical set up of Section 2.2, the information given to subjects satisfies all four properties. Proposition 3.1 in Klibanoff *et al.* (2012) shows how the predictions of the α-MEU model and the smooth ambiguity model differ when Properties One through Four are assumed. [38, 39] Preferences of the Schmeidler (1989) model under *uncertainty aversion* (i.e., with a convex *capacity* representation) coincide with those in an MEU model where the representation set of probabilities is the *core* of the convex capacity.[40] Hence, a Schmeidler model under uncertainty aversion whose convex capacity in the representation has a core that satisfies

[38] For the case of the smooth ambiguity model, Proposition 3.1 further assumes that the weights μ are uniform. This assumption is, of course, consistent with the information given to subjects under both 1-ball and 2-ball conditions. But the assumption is not necessary for the application to the theoretical setup of Section 2.2 provided $\mu(p_2) = \mu(p_3)$).

[39] As was noted in Footnote 2, the formal details of the set up in Klibanoff *et al.* (2012) are slightly different from what we have here. In particular, they have a product state space, unlike here. So, Property Four stated here does not apply literally to their setup. However, the belief about each component of the product space is unambiguous: it is a probability known to the DMs. Hence, the substantive element of Property Four is implicitly assumed in Proposition 3.1, even though the assumption is not explicitly stated in the proposition.

[40] A convex capacity is a set function $\nu: 2^S \to [0,1], \nu(E) \geq 0, \nu(S) = 1, \nu(E \cup F) \geq \nu(E) + \nu(F) - \nu(E \cap F)$. Every convex capacity ν has an non-empty *core*, a compact, convex set of probability measures defined as follows: $\mathcal{P}(\nu) \equiv Core(\nu) \equiv \{p \in \Delta(S) \mid p(E) \geq \nu(E), for\ all\ E \subset (S)\}$, where $\Delta(S)$ denotes the set of all probability measures on S. Furthermore, $\nu(E) = min_{p \in \mathcal{P}(\nu)} p(E)$. A *belief function* (also known as a *totally monotone capacity*) satisfies a stronger version of the third property specified for a convex capacity: for every $n > 0$ and every collection $E_1, \ldots, E_n \in 2^S$, $\nu(\cup_{i=1}^n E_i) \geq \sum_{I \subseteq \{1,\ldots,n\}, I \neq \emptyset} (-1)^{|I|+1} \nu(\cap_{i \in I} E_i)$, where $|I|$ denotes the cardinality of I.

Properties One through Four, will share the predictions of an α-MEU (with $\alpha = 1$) as specified in Proposition 3.1. In particular, this implies that there will be no strict preference for the mixed act over its constituents in the 2-ball condition. An example of a capacity whose core satisfies the four properties is a belief function $\nu: 2^S \to [0,1]$, satisfying the following further conditions: $\nu(Z_1) = r = \nu(Z_2), 0.2 \geq r > 0;\ \nu(\{Z_1, Z_2\}) = 0.5.$ [41]

However, under the 1-ball condition, the subject's information violates Property Three, since the information implies that $p(R_1) = p(B_1)$. Under such a restriction the ambiguity in each component resolves in an identical way implying mixing cannot in any way help with ambiguity hedging. As Klibanoff *et al*. (2012) point out, in this case the smooth model will, like the α-MEU model, also predict indifference to mixing.

Notice, given the set of probability distributions $\{p_1, \dots, p_4\}$ in Table 2, if a DM weights the distributions uniformly, the reduced probability measure on S under the 1-ball condition is the same as that under the 2-ball condition.[42] Thus, under the information available to the subjects, it is natural to expect that a probabilistically sophisticated DM (Machina and Schmeidler (1992)) will choose the same way under the 1-ball and 2-ball conditions. An example of a probabilistically sophisticated DM is a rank dependent utility maximizer à la Quiggin (1982). A rank dependent utility preference is another prominent member of the class of preferences representable as a Choquet expected utility (CEU). Such a preference is represented by a capacity obtained via a distortion of the probability measure on outcomes induced (jointly) by the given probability measure on S and an act. Hence, a subject in the experiment with such preferences, given the information available, should not choose differently under 1-ball and 2-ball conditions, since the reduced probability measure on states is identical under both conditions. Thus, these preferences do not predict a difference between the premia under the two conditions. In this sense, the prediction rank dependent utility is similar to that of an MEU preference rather than to a smooth ambiguity preference. Thus, two very prominent types of CEU preferences, the

[41] In Jaffray (1989), the model is a special case of α-MEU in that the set of probabilities is restricted to be given by the core of a belief function. In Jaffray's presentation, the set of probabilities represent objectively given imprecise information. Gul and Pesendorfer (2014) axiomatize a fully subjective, "Savage-style" version of Jaffray's "von Neumann-Morgenstern style" model. Olszewski (2007) posits and axiomatizes an α-MEU style model with an objectively given set of probabilities (on outcomes). Gajdos *et al*. (2008)'s "contraction" model, which falls in the MEU class, also takes the set of probabilities as objectively given to the DM. All these models share the prediction of the MEU model in our experiment.

[42] Actually, the assumption of a uniform distribution is not necessary for the conclusion in the context of our experiment. The reduced probability on suits is the same in the 1-ball and the 2-ball conditions, for any subject who obtains it by (standard) reduction from probability distributions over the possible compositions *and for whom the latter conform with the information given.* The significance of the italicized phrase is that, in order for the claim to hold, the DM must treat the draws from the opaque bag that determine, respectively, Deck 2, the Black-component of Deck 3, and the Red-component of Deck 3, as separate realizations of the same process. Given this, it does not matter whether the DM thinks "1" and "4" equally probable in that process.

uncertainty averse Schmeidler model and the rank dependent expected utility model, may be seen to share the predictions of the MEU model taking into account the information available to the subjects.[43]

Turning to the other side of the divide, we note first that the models of Ahn (2008), Ergin and Gul (2009), Nau (2006), Nielsen (2010) and Seo (2009) have substantial subcategories that share the same functional form representation with the smooth ambiguity model, and these must share the predictions of the smooth model in our main experiment. Second, a smooth ambiguity preference with the ambiguity attitude function $\phi(.)$ given by a negative exponential (constant absolute ambiguity aversion) is a Variational preference (Maccheroni *et al*. 2006). Third, given the set of probability distributions $\{p_1, \ldots, p_4\}$ in Table 2, if a DM weights the distributions uniformly, a smooth ambiguity preference is also a Vector Expected Utility (VEU) preference since the sufficient conditions noted in Table II (pp. 826) of Siniscalchi (2009) are then met. These points are illustrative. They are sufficient to show that our main experiment's results also support these models in so much as they support the smooth ambiguity model. But, of course, there may be other possibilities too (and, as discussed in Section 5, the equivalence does not in all cases carry over into our follow-up experiment).

Finally, it is apparent that our discussion of the divide between models exploits Properties One – Four. Thus, it does not rule out the possibility that the divide between models could be different if the beliefs of subjects about our 2-ball condition depart from those properties. Just as with our discrimination between α-MEU and smooth ambiguity models in the main text, our discrimination between broader classes of models is also contingent on grounded assumptions about beliefs.

[43] It is possible to have instances of CEU utility preferences that ignore part of the information available to the subjects and make predictions in our experiment akin to the smooth ambiguity model. Simon Grant gave us an example of such a convex capacity which violated Property Four and had choice implications like the smooth ambiguity model in the 2-ball condition case. Peter P. Wakker has also explained to us about these possibilities.

Appendix D: Summary statistics for main experiment

Table D1: Certainty Equivalents (in €, rounded to nearest cent)

	Ambiguity Seeking (n=16)			Ambiguity Neutral (n=50)			Ambiguity Averse (n=31)			All Subjects (n=97)		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Ambiguous acts												
$CE(f_1,$2-ball)	5.47	4.45	2.91	4.51	4.33	2.19	5.84	4.93	3.02	5.09	4.53	2.64
$CE(f_2,$2-ball)	5.28	4.50	3.01	4.65	4.15	2.74	6.03	4.98	3.24	5.19	4.48	2.98
$CE(mix,$2-ball)	5.67	5.38	3.56	4.87	4.48	2.81	6.99	5.68	3.99	5.68	4.98	3.45
$CE(g_1,$2-ball)	5.43	5.25	2.59	4.95	4.53	3.01	7.43	5.23	4.19	5.82	4.98	3.53
$CE(g_2,$2-ball)	6.35	6.25	3.31	4.95	4.53	3.01	5.53	4.53	2.82	5.37	4.98	3.02
Ambiguous acts												
$CE(f_1,$1-ball)	5.43	4.95	2.29	5.28	4.93	2.72	6.06	5.03	2.48	5.56	4.98	2.58
$CE(f_2,$1-ball)	5.75	5.45	2.20	5.45	4.98	2.68	6.04	5.53	2.29	5.69	5.03	2.47
$CE(mix,$1-ball)	6.27	5.73	3.30	5.45	4.33	3.60	6.17	5.03	3.10	5.81	4.88	3.38
Risky acts (denoted by chance of €20)												
$CE(0.7)$	11.65	11.33	3.30	10.58	11.18	3.80	11.53	12.78	3.00	11.06	11.73	3.48
$CE(0.3)$	5.59	5.25	2.00	6.19	5.73	3.05	6.39	5.73	2.82	6.15	5.58	2.81

Appendix E: Experimental Instructions

E1. *Supplementary information on procedures*

All subjects were recruited using an online recruitment system developed by Tilburg University. On average, sessions contained approximately 10 subjects and lasted for about 45 minutes. In the main experiment, 62 subjects participated under the tabular protocol and 35 under the verbal protocol. 86 subjects participated in the follow-up experiment. The instructions of the main experiment for the tabular protocol are given in Section E2. The instructions of the follow-up experiment are given in Section E3.

E2. *Instructions for the main experiment*

<The following was distributed on paper and read aloud.>

This experiment involves several decision tasks, each of which is a choice between two options involving amounts of money and/or chance. At the end of the experiment, the computer will select one choice to be for real for you. This means that you will be paid on the basis of what you chose in that choice, after any relevant chances have been resolved. In addition to the payment based on your responses to the choices, we will pay you €5 simply for participating.

Each choice is between two options. One of the options will be a certain amount of money whereas the other option will be a gamble. Here are some examples of gambles and an illustration of the kind of table with which your screen will display them:

	Spade	**Club**	**Heart**	**Diamond**
Gamble 1	€0	€0	€0	€20
Gamble 2	€20	€20	€0	€0
Gamble 3	€0	€0	Roll die is EVEN: €20 Roll die is ODD: €0	€0

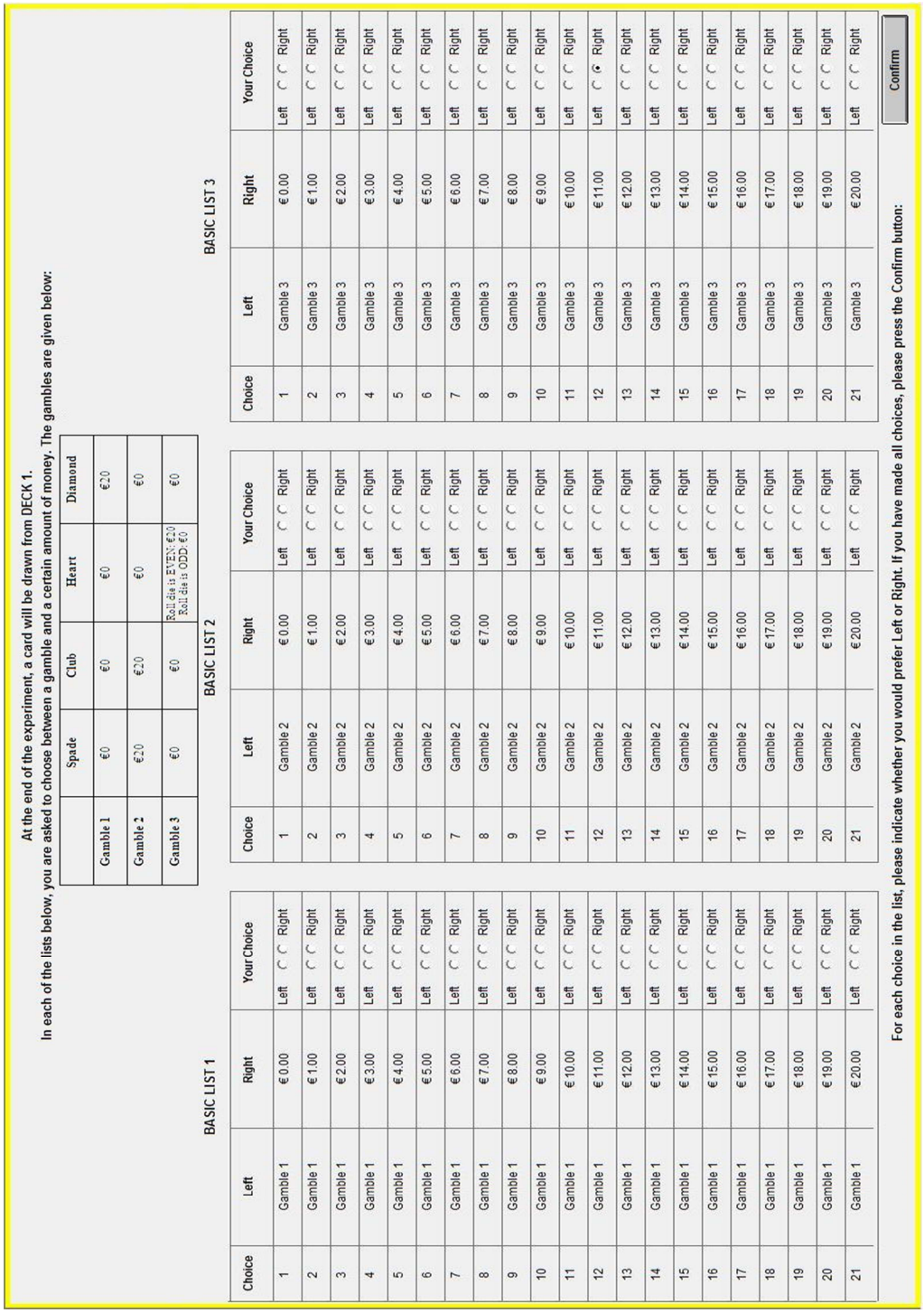
Looking at these examples, you can see that all gambles yield an outcome depending on the suit of a card drawn at the end of the experiment from a particular deck of cards. For example, Gamble 1 yields €20 if a diamond is drawn, and €0 otherwise. Similarly, Gamble 2 yields €20 if a spade or club is drawn, and €0 otherwise. As you can see, the outcome of Gamble 3 depends not only on the suit of a card drawn from a deck, but also on the roll of standard six-sided die performed at the end of the experiment. In this case, if a Spade, Club, or Diamond is drawn from the deck, Gamble 3 yields €0. If a heart is drawn from the deck, and the roll of the six sided die is even, Gamble 3 yields €20. If a heart is drawn, and the roll of the six sided die is odd, Gamble 3 yields €0.

The tasks will be grouped together in choice lists. Your computer will show you either 2 or 3 of these lists at once on a screen. An example screen is on the separate sheet provided. This example is not a screen that you will actually encounter during the experiment, but merely illustrates the general format. At the top of each screen, you can see a line of text that tells you that a card will be drawn at the end of the experiment from a particular deck, here Deck 1. During the experiment, you will encounter Decks 1, 2 and 3. I will tell you more about the decks later. For now, just note that the top of each screen will tell you the deck from which the card will be drawn to determine the outcome of the gambles described on that screen. Just below this, some gambles whose outcomes depend on the card drawn from the deck are presented in a table. In this example, there are three such gambles – the

same ones in the table above. In the bottom half of the screen, you can see three lists, labelled Basic List 1, Basic List 2 and Basic List 3. Each of them is a numbered list of 21 choices with the same basic format, which I will now explain by talking you through Basic List 1 on the example screen.

Each numbered row of List 1 describes a choice between two options labelled Left and Right. In each case, Left corresponds to choosing Gamble 1 described at the top of the screen, whereas Right corresponds to choosing a certain amount of money. Note that as you move down List 1, Left always corresponds to the gamble, whereas the amount of money offered by Right rises as you move down the list, starting at €0 and rising in increments of €1 to €20. This basic structure will be the same for every list that you encounter.

Figure E.1: Example Choice List



Note that Lists 2 and 3 are exactly the same, except that Left corresponds to choosing Gamble 2 in List 2 and in List 3, Left corresponds to choosing Gamble 3. The way you complete the lists is to select either Left or Right for each choice task on each list by using your mouse to click in the corresponding circle on that row. *You cannot proceed to the next screen until you have selected either Left or Right in each row of each list*. Now look closer at choice 1 in List 1. We imagine that all of you will wish to choose Left in Choice 1. This is because Left offers a chance of €20, whereas Right in Choice 1 yields a certain €0. We also imagine that, as the certain amount of money offered by Right increases, at some point you will wish to switch to Right. For example, even if you have not done so in an earlier choice, we imagine that you will wish to choose Right in Choice 21 because, in that choice, Right yields a certain €20, whereas Left only offers a chance of €20. A similar logic applies to each list on the screen. Thus, for each list, we imagine that all of you will wish to select Left in the first choice and perhaps in some further choices until, at some point, you will switch to Right and then stick with Right until the final choice in the list. In fact*, the computer will only accept choices that do have this form of just switching once from Left to Right in each list.* The reason for this is that we assume throughout that, if you choose the gamble offered by Left in some choice over a particular sum of money, then you would also wish to choose that gamble over any smaller sum of money, and, similarly, if you would choose a particular sum of money over the gamble, you would also choose any larger sum of money over that gamble. However, note that it is entirely up to you at which point in a choice list to make your switch from selecting Left to selecting Right. Just recall that each choice in the list could prove to be the one selected at the end to determine your payment. In that case, the choice you have made in it, together with any card draws or die roll, will determine the payment you receive. After completing all the lists on a screen, you press the Confirm button. When you press the Confirm button, the computer will check that you have made a selection in each choice on the decision screen and that, in each list, you have switched exactly once from Left to Right, as you move down the list. If this is not the case, the computer will prompt you to make alterations. But, if it is the case, the computer will accept your confirmed selections. You cannot make any changes to them after that. Note that we will wait for everybody to complete each list before proceeding; you might have to wait a while before a new list appears on your screen.

In the experiment, there are two kinds of lists, called "basic" and "zoomed-in." At the top of each list, it will say which list it is. Basic lists have the form described above. When you have completed such a list, you will have switched from Left to Right at some point. The computer will then show you a "zoomed-in" list. This is generated from the basic list by zooming-in on that part of the basic list where you switched from Left to Right. Thus, the gambles listed at the top of the screen and the deck from which a card will be drawn are the same as on the previous screen. However, in the zoomed-in list, the top choice will be the last choice in which you selected Left in the corresponding basic list, and the bottom choice will be the first choice in which you selected Right. These choices will be filled in for you, because you have already confirmed them. The computer then requires you to make further choices in which the sums of money offered by Right are intermediate between those on the top and bottom rows and rise as you move down the list, but now in increments of 5 cents. Again, you must make a selection for each choice of each list, and are only allowed to switch just once from Left to Right in the zoomed-in lists as well.

Each screen will have the format described above. The gambles that depend on a draw of a card will be described at the top of the screen, together with the name of the deck from which a card will be drawn. Remember that these things can differ from one screen to another so you should always study them carefully before completing each choice list. So far, I have not said much about the decks. This is because, at each point in the experiment, I will project on the whiteboard here some important information about the deck that the decisions you are currently taking relate to. At the end of the experiment, I will construct the decks in the ways specified by the information I will project on the board. Then, each deck will be shuffled, and a 10-sided die with numbers 1-10 will be rolled by one of you. Then, I will draw the card whose position from the top is given by the number on the die. For example, if the die roll is a 3, I will draw the 3rd card from the deck. This card will determine the outcome of any gamble that depends on a draw from that deck.

After I have done this for each deck, I will roll the 6-sided die to determine the outcome of any gamble that depends on this roll. Finally, your computer will select one choice at random from all of those in all the basic lists that you have completed. If this choice is not either the last choice in the list in which you chose Left or the first in which you chose Right, then what you chose in that choice task will determine your payment. On the other hand, if the randomly selected choice is either the last choice in the list in which you chose Left or the first in which you chose Right, the computer will move to the corresponding list on the relevant zoomed-in screen and randomly select one choice from that zoomed-in list. What you chose in that task will then determine your payment. Note that this simply means that every choice that you face could prove to be the one selected to be paid for real. If you chose Right in that choice, you will receive the amount of money specified by Right. If you chose Left, you will receive either €20 or €0, depending on the card drawn from the relevant deck and/or on the roll of the six-sided die.

<the following was read aloud only>

Each deck that we use in the experiment will have ten cards. First, I will ask you to complete some tasks relating to Deck 1. This slide gives you some information about Deck 1. It says that Deck 1 has seven spades and three hearts. At the end of the experiment, I will construct Deck 1 to contain ten cards in total, made up of seven spades and three hearts. Then I will draw a card from Deck 1, as explained in the instructions. Your computer will now display several lists involving Deck 1. Please complete them at your own pace.

<subjects complete lists for Deck 1>

Next, I will ask you to complete some tasks relating to Deck 2. This slide gives you some information about Deck 2. It says that Deck 2 has 5 black cards (i.e. either spade or club) and 5 red cards (i.e. either heart or diamond). However, it does not tell you exactly how many cards of each suit there will be in Deck 2. The way this will be resolved is as follows. I have here an opaque bag containing balls, all of which are numbered either 1 or 4. Here is an example of each. I am not going to tell you how many balls of each kind there are. At the end of the experiment, I will shake the bag and draw one ball from it. The number on this ball will give *both* the number of spades *and* the number of hearts in Deck 2. The numbers of clubs and diamonds will adjust accordingly, to make up 5 black and 5 red cards in total. Thus, as this slide shows, there are two possibilities for the composition of Deck 2, depending on the ball that is drawn. If that ball is numbered 1, the deck will contain 1 spade, 4 clubs, 1 heart and 4 diamonds. But, if the ball is numbered 4, the deck will contain 4 spades, 1 club, 4 hearts and 1 diamond. In both cases, there are 5 black cards and 5 red cards in total; and the number of spades equals the number of hearts. Note that how likely each of the two possible compositions of Deck 2 is depends on the contents of the bag of balls, which I have not revealed. The two compositions may or may not be equally likely. Your computer will now display several lists involving cards drawn from Deck 2. Please complete them at your own pace.

<subjects complete lists for Deck 2>

Next, I will ask you to complete some tasks relating to Deck 3. This slide gives you some information about Deck 3. Thus, as with Deck 2, Deck 3 has 5 black cards and 5 red cards, but the slide does not tell you how many cards of each suit there will be in Deck 3. For Deck 3, this will also be resolved in a similar way to for Deck 2, except that this time two balls will be drawn from the bag rather than one. I will shake the bag and draw one ball from it; then I will show you the number of this ball and place it back, before shaking the bag again and drawing another ball. The number on the *first* ball will give the number of spades in Deck 3 and the number of the *second* ball will give the number of hearts. As before, the numbers of clubs and diamonds will adjust accordingly to make up 5 black and 5 red cards in total. Thus, as the slide shows, there are four possibilities for the composition of Deck 3. If both balls drawn are numbered 1, Deck 3 will contain 1 spade, 4 clubs, 1 heart, and 4 diamonds. If the first ball drawn is numbered 1 and the second ball numbered 4, Deck 3 will contain 1 spade, 4 clubs, 4 hearts, and 1 diamond. If the first ball drawn is numbered 4 and the second ball numbered 1, the Deck 3 will

contain 4 spades, 1 club, 1 heart, and 4 diamonds. If both balls drawn are numbered 4, the Deck 3 will contain 4 spades, 1 club, 4 hearts, and 1 diamond. In all cases, there are 5 black cards and 5 red cards in total; the composition of the black cards is determined separately from the composition of the red cards, because the former is determined by the first ball and the latter by the second ball. Thus, unlike Deck 2, the number of spades and hearts may or may not be equal in Deck 3. Note that how likely each of the four possible compositions of Deck 3 is depends on the contents of the bag of balls, which I have not revealed. The four compositions may or may not all be equally likely. Your computer will now display several lists involving cards drawn from Deck 3. Please complete them at your own pace.

E3. *Instructions for follow-up experiment*

<Instructions were the same as those given in Appendix E2, except for the final two paragraphs. These are given below with differences marked in red. >

Next, I will ask you to complete some tasks relating to Deck 2. This slide gives you some information about Deck 2. It says that Deck 2 has 5 black cards (i.e. either spade or club) and 5 red cards (i.e. either heart or diamond). However, it does not tell you exactly how many cards of each suit there will be in Deck 2. The way this will be resolved is as follows. I have here an opaque bag. As you see, I am now going to fill the opaque bag with balls, all of which are numbered either 1 or 4. Here is an example of each type. I am going to put 10 balls in the bag. Before I do so, I would now like a volunteer to confirm to you that there is nothing else in the bag and that I am putting into it 5 balls of each type. At the end of the experiment, I will shake the bag and draw one ball from it. The number on this ball will give *both* the number of spades *and* the number of hearts in Deck 2. The numbers of clubs and diamonds will adjust accordingly, to make up 5 black and 5 red cards in total. Thus, as this slide shows, there are two possibilities for the composition of Deck 2, depending on the ball that is drawn. If that ball is numbered 1, the deck will contain 1 spade, 4 clubs, 1 heart and 4 diamonds. But, if the ball is numbered 4, the deck will contain 4 spades, 1 club, 4 hearts and 1 diamond. In both cases, there are 5 black cards and 5 red cards in total; and the number of spades equals the number of hearts. Note that each of the two possible compositions of Deck 2 is equally likely to be the one used, because the bag contains equal numbers of each type of ball. Your computer will now display several lists involving cards drawn from Deck 2. Please complete them at your own pace.

<subjects complete lists for Deck 2>

Next, I will ask you to complete some tasks relating to Deck 3. This slide gives you some information about Deck 3. Thus, as with Deck 2, Deck 3 has 5 black cards and 5 red cards, but the slide does not tell you how many cards of each suit there will be in Deck 3. For Deck 3, this will also be resolved in a similar way to for Deck 2, except that this time two balls will be drawn from the bag rather than one. At the start of the procedure for Deck 3, the contents of the bag will be exactly as they were for the case of Deck 2, because I will replace the ball that determined Deck 2. However, the procedure from that point on will be different for Deck 3, in the way just indicated. I will shake the bag and draw one ball from it; then I will show you the number of this ball and place it back, before shaking the bag again and drawing another ball. The number on the *first* ball will give the number of spades in Deck 3 and the number of the *second* ball will give the number of hearts. As before, the numbers of clubs and diamonds will adjust accordingly to make up 5 black and 5 red cards in total. Thus, as the slide shows, there are four possibilities for the composition of Deck 3. If both balls drawn are numbered 1, Deck 3 will contain 1 spade, 4 clubs, 1 heart, and 4 diamonds. If the first ball drawn is numbered 1 and the second ball numbered 4, Deck 3 will contain 1 spade, 4 clubs, 4 hearts, and 1 diamond. If the first ball drawn is numbered 4 and the second ball numbered 1, the Deck 3 will contain 4 spades, 1 club, 1 heart, and 4 diamonds. If both balls drawn are numbered 4, the Deck 3 will contain 4 spades, 1 club, 4 hearts, and 1 diamond. In all cases, there are 5 black cards and 5 red cards in total; the composition of the black cards is determined separately from the composition of the red cards, because the former is determined

by the first ball and the latter by the second ball. Thus, unlike Deck 2, the number of spades and hearts may or may not be equal in Deck 3. Note that each of the four possible compositions of Deck 3 is equally likely to be the one used, because the bag contains equal numbers of each type of ball. Your computer will now display several lists involving cards drawn from Deck 3. Please complete them at your own pace.