No Two Experiments Are Identical*

Larry G. Epstein Yoram Halevy

February 22, 2014

Abstract

We study choice between bets on the colors of two balls, where one ball is drawn from each of two urns. Though you are told the same about each urn, you are told very little, so that you are not given any reason to be certain that the compositions are identical. We identify choices that reveal an aversion to ambiguity about the relation between urns, thus identifying a source of uncertainty different from the usual Knightian distinction between risk and ambiguity. Choice behavior is studied in a controlled high-stakes laboratory experiment, and the ability of new and existing models to rationalize the experimental findings is examined.

*Boston University, lepstein@bu.edu and University of British Columbia, yoram.halevy@ubc.ca. We thank Kfir Eliaz, Faruk Gul, and Mike Peters for comments; discussions with Kyoungwon Seo and Chew Soo Hong were invaluable. Comments from audiences at Boston University, UBC, Northwestern, CETC 2013 in McGill (Montreal), D-TEA 2013, Decision Making under Uncertainty and Beyond (HEC, Paris), CIRPEE-WZB (Berlin) Workshop on Ambiguity in Experimental Markets, SITE (experimental economics section), ESA Meetings in Santa Cruz, 2014 North American Winter Meetings of the Econometric Society and 2014 Santa Barbara Conference on Experimental and Behavioral Economics are gratefully acknowledged. We are grateful also for the financial support of the National Science Foundation award SES-1216339 (Epstein) and the Social Science and Humanities Research Council of Canada (Halevy). David Freeman, Terri Kneeland and students from UBC's ECON490 class of 2013 (Experimental Economics section) provided excellent research assistance.

1 Introduction

We study choice between bets on the colors of two balls, where one ball is drawn from each of two urns. Though you are told the same about each urn, you are told very little; for example, you may be told only that each urn contains ten balls that are either red or blue. Accordingly, you are not given any reason to be certain that the compositions are identical, nor are you given any reason for being confident that the urn compositions are unrelated or related in any particular way. The following four questions are addressed. Is there behavior that intuitively reveals such a lack of confidence? Do subjects exhibit such behavior in a laboratory setting? How is the noted behavior associated with aversion to ambiguity as demonstrated by the two-urn Ellsberg experiment? How can the behavior that we identify be modeled?

The relation to the familiar two-urn Ellsberg experiment provides perspective. There you are given different information about the two urns: you are told the exact color composition for one, the risky urn, and nothing at all about the composition of the other one, the ambiguous urn. Then you are asked to choose between betting on a single draw from the risky urn or from the ambiguous urn. It has been shown that for many people the lack of information about the ambiguous urn is reflected in choice between such bets. Here, in contrast, the information about each urn's composition is imprecise and it is symmetric in the two urns, that is, both urns are ambiguous and they are indistinguishable (though not necessarily identical). We consider bets that pay according to the colors of balls drawn from both urns, and we argue that the choice between different bets of this sort reveals the subjects' attitude towards another dimension in which information is lacking–the relation between urns' compositions. Thus just as Ellsberg gives behavioral meaning to one kind of ambiguity and provides a litmus test for related models, we strive to do the same for a setting with "repeated random events"–aversion to ambiguity about heterogeneity.

Below we often refer to "repeated experiments" rather than to "repeated random events." Thus "experiment" is used in two senses: either as the realization of a stock return or another economically relevant random variable, or in reference to a laboratory experiment on choice behavior. However, the meaning should be clear from the context; in particular, we will never need to refer to repeated laboratory experiments. Note also that just as the draws from the two urns are made simultaneously, repeated experiments should be understood to be cross-sectional rather than ordered in time.

The choice behavior that we identify is then tested in a controlled laboratory experiment where high monetary stakes are used to incentivize subjects' choices between bets. The findings suggest that many subjects are averse to the lack of information concerning the relation between the urns' compositions, and that this aversion is associated with (but distinct from!) Knightian ambiguity aversion as measured in the standard Ellsberg experiment. They also reflect on, and discriminate between, some preference models that have been studied in the literature on repeated experiments.

The most popular model of preference in a setting with repeated experiments is the *exchangeable Bayesian model,* which is the specialization of subjective expected utility due to de Finetti (1937). The model has three important (indeed, defining) features. First, it implies indifference between any two bets on the outcomes of experiments that differ only in a reordering of experiments; refer to such indifference as symmetry, or, following de Finetti, as exchangeability. Such symmetry is intuitive in the case of our urns where the information given provides no reason for distinguishing between them, and it is natural more generally as illustrated in the less contrived examples described shortly. A second feature is that preference over bets (or acts) is probabilistically sophisticated (Machina and Schmeidler, 1992), which means, roughly speaking, that beliefs are probabilistic (can be represented by a probability measure). Probabilistic sophistication does not require the expected utility functional form, but the latter is the third noteworthy component of the exchangeable Bayesian model. Not surprisingly in light of the literature surrounding the Ellsberg paradox, the latter model is inconsistent with the behavior that we identify as revealing aversion to ambiguity about how the urns differ. Indeed, the contradiction is more basic because the behavior contradicts probabilistic sophistication even without the expected utility functional form. For all of the above reasons, in our discussion of models (Section 4) we explore first generalizations of the de Finetti model that retain exchangeability (or symmetry) but not probabilistic sophistication.

We show that one of these generalized models can account for the marginal distribution of observed choices. However, they struggle with the associations present in the data. This motivates us to outline (see Section 4.2) an alternative to the de Finetti-style model that centers on multiple "sources" (Tversky and Fox, 1995; Tversky and Wakker, 1995) or "issues" and that translates objects of choice into multistage lotteries which are evaluated recursively (see, for example, Segal (1987), Ergin and Gul (2009), Amarante, Halevy and Ozdenoren (2013)). In our case, there are three issues–risk, bias (the composition of each urn), and uncertain differences between urns–and accordingly objects of choice are translated into three-stage lotteries. The resulting model retains a Bayesian prior over each source, yet has sufficient flexibility to rationalize the observed behavior, overcoming the strict restrictions imposed on the associations by the various generalizations of the de Finetti model.

1.1 Economic significance

Betting on the draws from a sequence of urns is intended as a canonical example of choice problems where payoffs to an action depend on the realization of multiple random events. Suppose, for example, that the outcome s_i of the i-th experiment is given by an equation of the form

$$s_i = \beta \cdot x_i + \epsilon_i,\ i = 1, 2, ..., I. \tag{1.1}$$

Experiments may differ and the vectors x_i describe the observable heterogeneity.[1] The key issue is the decision maker's model of the residuals or unobserved heterogeneity ϵ_i, which are the source of the uncertainty she faces. If all sources of heterogeneity of which she is aware are included in the x_is, then it is natural that she be indifferent between any two bets on the realization of residuals that differ only in a reordering of experiments. However, the individual may not be confident that the x_is describe all relevant differences between experiments, in which case she may not be certain that residuals are identical, or that they are related in any particular way. Though she may not be able to describe further forms of heterogeneity, she may be worried that there are gaps in her understanding that could be important and thus she may wish to take into account their possible existence when making choices.

A number of studies have argued for the importance of the noted lack of confidence. They serve also to illustrate decision making with repeated experiments in economic settings; in the first three examples, the decision maker can be thought of as a policy maker. In the context of the cross-country growth literature where an experiment corresponds to a country and the outcome is its growth rate, Brock and Durlauf (2001) point to the open-endedness of growth theories as a reason for skepticism that all possible

[1] In the urns context, there is no observable heterogeneity

differences between countries can be accounted for (p. 231), and they emphasize the importance of "heterogeneity uncertainty." King (2001) makes a similar critique in an international relations context where an experiment corresponds to a pair of countries and the outcome is conflict or lack of conflict; he refers to "unmeasured heterogeneity."[2] This paper complements these critiques by translating them into behavioral terms and thus giving more precise meaning to a concern with "heterogeneity uncertainty" or "unmeasured heterogeneity."

The applied IO literature provides an example of a different sort.[3] Here there is a cross section of markets in each of which an entry game is played. Thus an experiment is a market and an outcome is the number and identity of entrants in a pure strategy Nash equilibrium. The difficulty faced by the policy maker is that there may be multiple equilibria and she has little understanding of how equilibria are selected, and accordingly how selection mechanisms may differ or be related across markets. A fourth example arises in repeated English auctions when, as in Haile and Tamer (2003), because of the free-form nature of most English auctions in practice, one makes weak assumptions about bidders' behavior. Then equilibrium behavior in each auction is multiple-valued and can be narrowed down and related across auctions only via heroic and often unjustifiable assumptions. This has implications for an auctioneer who is choosing reserve prices (Aryal and Kim, 2013). Though our laboratory experiment does not investigate behavior in the above specific settings, the results lend support to the hypothesis that decision-makers care about poorly understood differences across markets or auctions.

Finally, as a motivating example involving a more standard individual decision problem, consider the problem of optimal (static) portfolio choice. There are I securities available and the individual's model of returns is a linear factor model as in arbitrage pricing theory (APT). That is, the ith return s_i is as in (1.1), which in more natural notation takes the form

$$s_i = \beta_i \cdot X + \epsilon_i,\ i = 1, 2, ..., I;$$

the vector X gives factor returns and β_i gives the betas or factor loadings

[2] He writes (p. 498) that "in international conflict data are neither powerful nor even adequate summaries of our qualitative knowledge", so that the common assumption of exchangeability is usually violated.

[3] We are referring to the literature on entry games and partial identification (see Tamer (2010) and the references therein). For an explicit choice-theoretic perspective, see Epstein and Seo (2013).

of security i. APT typically adopts strong assumptions on the idiosyncratic terms ϵ_i, say that they are i.i.d., which are suspect intuitively as explained in connection with (1.1). It is well known that ambiguity about returns can have substantial effects on the nature of optimal portfolios (Garlappi, Uppal and Wang, 2007), on participation decisions (Dow and Werlang, 1992), and on equilibrium pricing and uncertainty premia. In the presence of multiple securities, ambiguity implies that gains from diversification are limited and, on the pricing side, that idiosyncratic uncertainty can have a positive price in equilibrium even in the limit as I goes to infinity (Epstein and Schneider, 2010). These latter implications are contrary to the intuition based on the classic Law of Large Numbers (LLN). However, the latter does not apply when there is ambiguity about heterogeneity of security returns (that is, of the ϵ_is) of the sort studied here. Thus our experimental findings are relevant (albeit indirectly) to important issues in the theories of portfolio choice and asset pricing.

2 A Thought Experiment

Two urns, numbered 1 and 2, each contain 10 red or blue balls. A ball is to be drawn from each urn simultaneously. Beforehand, a subject is asked to choose between specified bets (or Savage acts depending) on the colors of the two balls. The draw from urn i yields an outcome in $S_i = \{R_i, B_i\}$, $i = 1, 2$, and the two draws together yield an outcome in

$$S_1 \times S_2 = \{R_1B_2, B_1R_2, R_1R_2, B_1B_2\}.$$

Denote by R_1B_2 both the obvious event and the corresponding bet that yields the prize 100 if that event is realized and the prize 0 otherwise; similarly for other events and bets. Prizes are denominated in dollars. Both the event and the bet $\{R_1B_2, R_1R_2\}$ are sometimes denoted simply R_1. More generally, the individual is asked to choose between acts over the state space $S_1 \times S_2$ having dollar outcomes. Any such act can be represented in the form

$$f = \begin{bmatrix} f(R_1B_2) & R_1B_2 \\ f(B_1R_2) & B_1R_2 \\ f(R_1R_2) & R_1R_2 \\ f(B_1B_2) & B_1B_2 \end{bmatrix}$$

Let $\succeq$ be a preference relation on the set of acts. We assume symmetry of information about the two urns, which in behavioral terms means that the individual is indifferent between any f and the act obtained by switching the prizes on R_1B_2 and B_1R_2.[4] Two immediate implications are that the individual is indifferent between the bets R_1B_2 and B_1R_2, and also between the bets R_1 and R_2 and between the bets B_1 and B_2. We interpret this indifference as reflecting the lack of observable differences between urns.

We assume further that the description of urns is symmetric in colors and thus that there is indifference between f and the act obtained if the prizes on R_1R_2 and B_1B_2 are reversed. This assumption is made solely for concreteness and simplicity. Together, the two symmetry assumptions imply

$$R_1B_2 \sim B_1R_2,\ R_i \sim B_i,\ i = 1, 2, \text{ and } R_1R_2 \sim B_1B_2. \tag{2.1}$$

We restrict attention throughout to preferences satisfying these indifferences, even where not stated explicitly.

Consider two choice problems and behaviors for this setting. The first, that we term *One vs Two*, offers the individual the choice between betting on the color drawn from one urn as opposed to betting on the colors drawn from both urns. More precisely, denote by *Same* and $Diff$ the bets that the two draws yield the same color and different colors respectively. (Thus $Same = \{R_1R_2, B_1B_2\}$ and $Diff = \{R_1B_2, B_1R_2\}$.) Then consider the choice between R_1 and *Same*, and also between R_1 and $Diff$. Consider in particular the following rankings, abbreviated below by One $\succ$ Two:

$$R_1 \succ Same \text{ and } R_1 \succ Diff. \tag{2.2}$$

The first thing to note about these rankings is that they contradict probabilistic sophistication. To see this, if P is any predictive prior on the state space $\{R_1B_2, B_1R_2, R_1R_2, B_1B_2\}$ representing beliefs, then the two rankings imply

$$P(R_1) > P(\{R_1R_2, B_1B_2\}) \text{ and } P(R_1) > P(\{R_1B_2, B_1R_2\}).$$

But $P(R_1) = P(B_1)$ by the symmetry in (2.1). It follows that R_1 has higher probability than *Same* and that the complement of R_1 has higher probability than the complement of *Same*, which is impossible. A contradiction is obtained similarly if both strict rankings in (2.2) are reversed, or if there is indifference in exactly one of them.

[4] In the terminology of Chew and Sagi (2006), R_1B_2 and B_1R_2 are *exchangeable events*.

Remark 2.1 *We emphasize that the contradiction is with the reliance on any single prior rather than with a specific prior. As a result, treating omitted variables as "nuisance parameters" and integrating them out does not resolve the conflict unless this is done in such a way as to violate probabilistic sophistication. The parallel with the Ellsberg paradox is that if uncertainty about the true composition of an urn is modeled solely via a prior over probability laws, then, as Savage noted, merely integrating over this uncertainty leaves the individual with the "mean" predictive prior, and still precludes the intuitive ambiguity averse behavior pointed to by Ellsberg.*

Given that (2.2) contradicts probabilistic sophistication, why would an individual make these choices? The intuition is that only the bets on both draws are subject to ambiguity about how urns differ or are related, which may, depending on the degree of aversion to such ambiguity, lead to the preference for R_1. To elaborate, consider the bet $Same$. This is an attractive bet if it is believed that the compositions of the two urns are similar, which would make "positive correlation" between draws likely. Since you are not told anything to the contrary, this belief is plausible but no more so than the belief that the two compositions are different–one urn is biased towards red and the other towards blue–which would make "negative correlation" between draws more likely and render $Same$ an unattractive bet. Given a conservative attitude, this uncertainty would act against choosing $Same$. Of course, there is also reason for a conservative individual to discount the bet R_1 because the composition of each urn is ambiguous. Therefore, speaking informally, we interpret the preference for R_1 as indicating a greater aversion to ambiguity about differences between urns than to ambiguity about the bias of any single urn. Similarly for the interpretation of the preference for R_1 over $Diff$. When both rankings in (2.2) are reversed strictly, (as for the model in Appendix B), a greater aversion to ambiguity about bias is indicated.

An alternative response to the problem of One vs Two is to choose

$$\text{either } Same \succeq R_1 \succeq Diff, \text{ or } Diff \succeq R_1 \succeq Same. \tag{2.3}$$

These choices do not indicate an aversion to ambiguity because both cases are consistent with probabilistic sophistication. For example, the first is rationalized by any probability measure satisfying

$$P(B_1 B_2) = P(R_1 R_2) \geq P(R_1 B_2) = P(B_1 R_2).$$

A Bayesian with an i.i.d. prior uniform within each urn would be indifferent between all three bets indicated.

Next we describe another choice, this time between nonbinary acts, that violates Savage's Sure-Thing-Principle (STP) and that can be understood as revealing an aversion to ambiguity about heterogeneity. Consider the following choice pattern that we term the *Correlation Certainty Effect (CCE)*:

$$f_0 \equiv \begin{bmatrix} 100 & R_1B_2 \\ 0 & B_1R_2 \\ 0 & R_1R_2 \\ 0 & B_1B_2 \end{bmatrix} \sim \begin{bmatrix} x & R_1B_2 \\ x & B_1R_2 \\ 0 & R_1R_2 \\ 0 & B_1B_2 \end{bmatrix} \equiv g_0 \text{ and} \tag{2.4}$$

$$f_1 \equiv \begin{bmatrix} 100 & R_1B_2 \\ 0 & B_1R_2 \\ x & R_1R_2 \\ x & B_1B_2 \end{bmatrix} \prec \begin{bmatrix} x & R_1B_2 \\ x & B_1R_2 \\ x & R_1R_2 \\ x & B_1B_2 \end{bmatrix} \equiv g_1 \tag{2.5}$$

The indifference $f_0 \sim g_0$ indicates that x is a conditional certainty equivalent for the bet on R_1B_2, where conditioning is on the two draws yielding different colors. Because the pair f_1 and g_1 is obtained from f_0 and g_0 by a change in common outcomes, (from 0 to x on the event $\{R_1R_2, B_1B_2\}$), the STP would require that f_1 and g_1 be indifferent. However, there is intuition that aversion to ambiguity about heterogeneity can lead to g_1 being strictly preferable. For the indifference $f_0 \sim g_0$ to obtain the individual might require a large value of x to compensate for the fact that the event where different colors are drawn is ambiguous. However, that ambiguity is completely eliminated when outcomes are changed as indicated which means that the individual is left with what now seems like an exceedingly large constant payoff. Put another way, ambiguity about the correlation between urns means that there is "complementarity" between what happens on $\{R_1B_2, B_1R_2\}$ and on its complement, contrary to the weak separability required by STP. The change in common outcomes also improves f_1 relative to f_0 but the effect is plausibly smaller there.

There is an alternative interpretation of CCE that is unrelated to ambiguity. The individual could be probabilistically sophisticated but, after using her predictive prior to translate acts into lotteries, she does not use vNM expected utility theory to evaluate the induced lotteries; for example, she may behave as in the Allais paradox. The experimental design (Section

3.1) permits us to separate between the two possible interpretations at the individual level. Since choices made in the Ellsberg problem and One vs Two could unambiguously indicate lack of probabilistic sophistication, we will use the choices made in these two problems to suggest whether a subject who exhibits CCE does so due to ambiguity or Allais-type behavior. Hence, if the subject is not probabilistically sophisticated in at least one of these choice problems, (for example, if she exhibits $One \succ Two$), then we are on stronger ground in interpreting CCE as being due to her aversion to ambiguity about the relation between the two urns.

We turn now to describing our experimental investigation of One vs Two and CCE.

3 A Laboratory Experiment

3.1 Design

Subjects were recruited from UBC's Vancouver School of Economics subject pool using ORSEE (Greiner, 2003) to an experiment that promised participants a chance to earn up to \$111 during a one hour experiment in decision making (including a show-up fee of \$10). After consent forms were signed, the instructions were read aloud. Subjects were presented with two urns which contained ten red or blue balls, (the language in the experiment used jars and marbles that were blue or green), and then they were asked to make binary choices between bets in each of ten questions (or choice problems). Complete instructions may be found in Appendix E.

The first eight questions were organized in pairs, which allowed us to infer strict preference from choices by slightly varying the prizes. For example, the first question asked the subject to choose between a bet paying \$100 if the ball drawn from urn 1 is red and a bet paying \$101 if the balls drawn from the two urns are of the same color. Question 2 was similar except that the two winning prizes were switched (\$101 if red is drawn from urn 1 and \$100 if the same color is drawn from both urns). The choice to bet on the single urn in question 1 implies a strict preference also when the two prizes are equal; and similarly if the choice in question 2 is to bet on the two colors being the same. Choice of both bets that pay \$100 is inconsistent with monotone preferences (assuming transitivity). Choice of both bets that pay \$101 is consistent with indifference between the bets. The rationale behind this design was explained to subjects before they answered any questions. The design establishes an upper (lower) bound on the indifference (strict preference) class.[5] In the description of results below, indifference should be interpreted as the absence of evidence for strict preference.

This design of questions, which we have not seen used previously, allows us to identify a strict ordinal ranking without using a cardinal valuation. Methods based on elicitation of cardinal valuations of bets, such as Becker-DeGroot-Marschak (1964) used in Halevy (2007) or a discrete version using a choice list used in Abdellaoui, Baillon, Placido and Wakker (2011), rely for incentive compatibility on separability of preferences, which often is not satisfied by non-expected utility models (Karni and Safra, 1987). In addition, cardinal elicitation is cognitively taxing, leading to possible errors in the elicitation procedure.

All subjects were presented with the following ten choice problems.

> *One* vs *Different*: Choose between a bet on a color of a ball drawn from a single urn and a bet that pays if the balls drawn from the two urns have different colors (questions 1 and 2).
>
> *One* vs *Same*: Choose between a bet on a color of a ball drawn from a single urn and a bet that pays if the balls drawn from the two urns have the same color (questions 3 and 4).
>
> *Same* vs *Different*: Choose between a bet that pays if the balls drawn from the two urns have the same color, or a bet that pays if they have different colors (questions 5 and 6).
>
> Standard Ellsberg: Questions 7 and 8 asked the subjects to choose between a bet on the color of the ball drawn from one of the two urns and a bet on a color of a ball drawn from a third urn that contained five red and five blue balls (a risky urn as in Ellsberg).
>
> Correlation Certainty Effect (CCE): In question 9 the subject was presented with a choice list in which she was asked to choose between a bet paying \$100 if the colors of the balls drawn are red from urn i and blue

[5] As noted above, the choice of the two bets paying \$101 in both questions is consistent with indifference. However, it could be that lowering the higher prize to $\$100 + \varepsilon$ (for example, \$100.1) in both questions might cause a subject to choose the same bet (with a prize of \$100 and $\$100 + \varepsilon$) in both questions, thus revealing strict preference.

from urn j (i and j, $i \neq j$, were chosen by the subject ex-ante), and \$$x$ if the two balls are of different colors. The choice list varied x between 1 and 100, permitting elicitation of an approximate conditional certainty equivalent. Denote by $\overline{x}$ the highest value of x for which the subject preferred the \$100 bet. After answering this question, $\overline{x}$ was inserted into question 10, which was not revealed to the subject beforehand, and the subject was asked to choose between receiving \$$\overline{x}$ for sure and a bet paying: \$100 (\$0) if the colors of the balls drawn from urns i and j are red and blue, and \$$\overline{x}$ if the two balls have the same color.

To sum up, questions (1-4) correspond to One vs Two, questions (9-10) correspond to CCE, and questions (7-8) correspond to the standard two-urn Ellsberg choice problem. Questions (5-6) elicit the preference between betting on the two draws yielding the same as opposed to different colors; as explained in the sequel, these questions permit identifying inconsistencies in answering questions (1-4).

Payment was determined by randomly choosing one question *before* subjects made their choices. This version of the Random Incentive System (RIS) is theoretically incentive compatible in eliciting ambiguity attitudes (Baillon, Halevy and Li, 2013), because the order suggests that choices are between lotteries over Savage acts wherein ambiguity cannot be hedged. In the standard implementation of the RIS (where the randomization is performed *after* choices have been made), the mechanism induces a choice problem in which the subject faces Anscombe-Aumann acts and thus where she can hedge ambiguity. It is important to note however that the randomization order in which the subject evaluates her options is determined cognitively and requires more empirical investigation.[6] The experimental implementation employed two subjects who wrote the random question's number on notes that were put into sealed envelopes and distributed among the subjects. The envelopes were opened only after all choices were made and the balls were drawn from the urns.

Before subjects made any payoff relevant choices, they were presented with four pairs of bets which demonstrated the natural symmetry between urns and between colors in this experiment. Then every subject was asked whether she agreed with the four indifferences in (2.1). This question was

[6] The empirical question whether subjects hedged in spite of our implementation is discussed further below in Section 3.2.5.

not incentivized.[7]

In order to eliminate a potential suspicion that the experimenter could manipulate the composition of the urns, each subject was asked to choose at the beginning of the experiment an urn (1 or 2) and a color (red or blue) to bet on in the bets that involve a single urn. Similarly, for the CCE questions (numbers 9 and 10), the subjects chose the urns to determine if she will be paid \$100 in R_1B_2 or B_1R_2. To simplify exposition of the results, we proceed as though all individuals chose urn 1 and the color red. Thus, for example, the bet on a single urn is represented by R_1, the bet on drawing red from urn 1.

3.2 Results

A total of 80 subjects participated in 4 sessions, which took approximately one hour each. Subjects were paid a total of CA\$4,851 (an average of just over \$60 per subject). Out of the 80 subjects, a total of 24 were removed from the analysis. 11 subjects were removed due to non-monotone choices (assuming transitivity, choosing in at least one pair of questions two lotteries with prizes of \$100), 7 due to non-transitive choices (assuming monotonicity, for example: revealing that $Same \prec R_1 \prec Diff$ in questions 1-4 but choosing consistently with $Same \succ Diff$ in questions 5-6), and 4 for disagreeing with the symmetry over colors and urns expressed in (2.1).[8] This leaves 56 subjects whose choices are analyzed below.[9] We consider this retention rate to be high when taking into account both the many dimensions along which choices are measured, and the strong consistency (transitivity and monotonicity) that we imposed.[10] We attribute this rate to the high stakes (more than \$100) employed in the experiment, which provided subjects sufficient incentive to

[7] For the most part our interpretations of experimental results presume the symmetry expressed in (2.1). However, as explained in Remark 3.1, the experimental design also permits (in principle) identification of failure of probabilistic sophistication without reliance on symmetry.

[8] Two mores subjects in the first session were caught cheating and their choices were excluded from the analysis (one of these subjects had non-transitive choices, so her/his answers would be removed in any case).

[9] As explained in the sequel, 7 more subjects are removed from the analysis of CCE.

[10] Two less stringent alternatives are to analyze aggregate responses, or to remove only problematic questions instead of omitting the problematic subject completely. However, our retained sample size is sufficiently large to give us very reliable answers about tendencies and associations in the population. See also Appendix D.

minimize arbitrariness and to consider their choices seriously.[11] Appendix D presents results with less strict retention criteria that include 77 subjects; the main findings reported below are found in the larger set as well.

Ellsbergian Ambiguity	**One vs Two**			**Total**
	One$\succ$Two	**One$\sim$Two**	**One$\not\succsim$Two**	
Averse	19	9	9	37
Neutral	3	10	2	15
Seeking	1	1	2	4
Total	23	20	13	56
Fisher exact test p-value=0.03<0.05				

Table 3.1: Ellsbergian ambguity and One vs Two

Table 3.1 displays the distribution of choices in the two problems that tested probabilistic sophistication: Ellsberg's two-urn classic problem and One vs Two. The notation One$\succ$Two denotes the two weak rankings in (2.2), when at least one ranking is strict. We write One$\not\succsim$Two if at least one of these weak rankings is violated, and One$\sim$Two if there is no strict preference between a bet on the color of a ball drawn from a single urn and the two bets ($Same$ and $Diff$) that depend on the colors of the balls drawn from the two urns. Table 3.1 is discussed further below.

3.2.1 Ellsbergian Ambiguity

Out of 56 subjects, two thirds (37 subjects) exhibited strict Ellsbergian ambiguity aversion when asked to choose between a bet on a color of a ball drawn from the risky urn (which contained 5 red and 5 blue balls) and a bet on a chosen color from one of the urns with unknown composition. Slightly more than a quarter of the subjects chose in a way that does not reveal ambiguity aversion or seeking and the remaining 4 subjects exhibited ambiguity seeking. That is, about 73% of the subjects are not probabilistically sophisticated in a standard Ellsberg experiment. This proportion is consistent with existing

[11] Subjects found the high stakes very motivating. As an illustration, two subjects in the first session changed the color of the ball they chose to bet on after the balls were drawn from the urns; both were excluded from the analysis (only one's choices were consistent elsewhere). We made sure that in later sessions subjects did not have such an opportunity.

studies that use certainty equivalent elicitation or choice data.[12,13]

3.2.2 One versus Two

There were 23 subjects (41% of 56 subjects) who exhibited One$\succ$Two, and 4 subjects out of the 13 classified under One$\not\succeq$Two exhibited Two$\succ$One (one of them was ambiguity averse and one was ambiguity seeking). These 27 subjects (48% of 56) violated probabilistic sophistication. The choices of the remaining 29 subjects (52% of the 56 subjects) can be rationalized by probabilistic beliefs. Out of the remaining 9 subjects classified under One$\not\succeq$Two, 2 subjects exhibited $Same \succ R_1 \succ Diff$ (one of them was ambiguity averse and one was ambiguity seeking), and 7 subjects exhibited $Diff \succ R_1 \succ Same$ (all exhibited Ellsbergian ambiguity aversion).

Combining both the Ellsberg and One vs Two choice problems, we find that only 10 out of the 56 subjects made choices that are consistent with probabilistic beliefs (see Table 3.1). That is, more than 82% were not probabilistically sophisticated in at least one of the choices. Moreover, the association between these two measures of attitude to uncertainty concerning different dimensions of the environment is highly significant (p-value Fisher exact test $0.03 < 0.05$). For example, out of 15 subjects that were neutral to ambiguity in the Ellsberg problem, 10 did not exhibit either One$\succ$Two or Two$\succ$One; and out of 23 subjects that exhibited One$\succ$Two, 19 were ambiguity averse in the Ellsberg problem.

Remark 3.1 *The above statements rely on the assumption of symmetry in urns and colors expressed in (2.1). However, the experimental design permits us to identify failure of probabilistic sophistication even without this assumption. For example, suppose that $R_1 \succ Same$ and $R_1 \succ Diff$. Under probabilistic sophistication, it would follow that the subjective probability of*

[12] Recent experimental studies that use probability equivalents find a much larger proportion of ambiguity neutral subjects. This substantial difference between elicitation modes deserves separate experimental and theoretical attention. We view binary choice data (questions 1-8 in the current study) as the experimental "gold standard", and we leave to proponents of the probability equivalent elicitation method to account for the difference.

[13] Note that this is an upper bound on the number of ambiguity neutral subjects; that is, the proportion of ambiguity neutral subjects may be even smaller since the increment of \$1 used in the experiment to detect strict preference may have been too big for some subjects.

drawing red is greater than 0.5, which would imply the preference to bet on drawing red from the ambiguous urn in the two-urn Ellsberg problem. Thus the choices indicated by the preceding rankings, together with the preference for betting on red in the risky urn rather than in the ambiguous urn, are inconsistent with probabilistic sophistication even without reliance on any symmetry assumption. Accordingly, the 19 subjects, more than one third of the total, who exhibited One$\succ$Two and simultaneously were ambiguity averse in the Ellsberg problem cannot be probabilistically sophisticated even if (2.1) is not satisfied.

One substantial difference between the two noted behaviors is that the proportion of subjects whose choices are consistent with the existence of probabilistic beliefs in One vs Two is almost double that found in the standard Ellsberg problem. This difference can be understood in light of the differences between the two choice problems. The choice of the risky urn in the classic Ellsberg problem leaves the subject with a purely risky bet in which the probability of winning is 50%. Thus the preference to bet on the risky urn, and hence violation of probabilistic sophistication, arises given only aversion to uncertainty about the bias of the ambiguous urn. In contrast, in the choice problem One vs Two all alternatives are ambiguous: the bet on a single urn is subject to ambiguity about the uncertain composition of the urn and the bets $Same$ and $Diff$ are ambiguous because of ambiguity about how urns differ or are related. The intuition for the preference to bet on one urn rather than on both is that the latter ambiguity is perceived to be more important. For some subjects ambiguity about bias could have been more important, and this might have led to the ranking Two$\succ$One which also contradicts probabilistic sophistication. However, there is another possibility–namely one of the rankings in (2.3). The bets $Same$ and $Diff$ could be viewed as relatively unambiguous because of a strong belief about how the urns were constructed. As an extreme example, suppose that the subject's hypothesis is that the experimenter drew two balls without replacement from an auxiliary urn containing one red and one blue ball, and if a red (blue) ball was drawn first, then urn 1 was filled with 10 red (blue) balls; the composition of urn 2 was determined in a similar fashion. Then it is certain that the balls drawn from urns 1 and 2 have different colors and thus $Diff$ pays 100 and $Same$ pays 0, each with certainty, so that R_1 is ranked between them.[14] The latter might arise also more generally from the feeling that

[14] Seven subjects exhibited $Diff \succ R_1 \succ Same$, and two exhibited the reverse rankings.

"there are only so many red balls to go around," say because the urns are thought to have been constructed by drawing without replacement from an auxiliary urn containing $n \geq 10$ balls of each color. The description of urns given to the subjects does not suggest this perception but there is no reason to rule it out. Probabilistic sophistication is consistent also with the ranking $Same \succ R_1 \succ Diff$, which might arise if urns are perceived to have a common component, so that a red draw from urn 1 indicates that a red draw is more likely also from urn 2. (The preference for $Same$ over $Diff$ is intuitive if the preceding construction is modified so that draws from the auxiliary urn are made with replacement; it is also an implication of exchangeable models as described in the next section.) The point is that there is more scope for probabilistically sophisticated behavior in One vs Two than in Ellsberg, even for a subject who dislikes (or alternatively likes) ambiguity. Given the symmetry conditions (2.1), violation of probabilistic sophistication involves two rankings (R_1 versus $Same$ and R_1 versus $Diff$), while in Ellsberg it depends only on the single comparison between R_1 and the bet on red from the risky urn.[15] Finally, even in the special case where $Same$ and $Diff$ are indifferent (consistent with the choices made by 27 subjects), it is important to note that such indifference does not imply that these bets are unambiguous. It could very well be that they are both very ambiguous and equally so with the bias of each urn, consistent with $R_1 \sim Same \sim Diff$ and hence also with probabilistic sophistication. The analogous Ellsberg-style experiment would have two urns, or sources, with each ambiguous and where their description would leave "equally ambiguous" at least plausible.

[15] There is a similarity with a modified Ellsberg experiment where information about the colors is *not* symmetric. Then an individual might choose to bet on one color from the ambiguous urn and on the other color from the risky urn. Though she may dislike ambiguity, these two binary choices are consistent with beliefs represented by an asymmetric single prior.

3.2.3 Correlation Certainty Effect

The analysis of CCE is based on 49 subjects.[16] As indicated in Table 3.2, 28 subjects (57%) exhibited the CCE. The remainder chose consistently with the Sure Thing Principle.[17]

	CCE	not **CCE**	**Total**
# of subjects	28	21	49
% of subjects	57.1%	42.9%	100

Table 3.2: Correlation Certainty Effect

As explained in Section 2, CCE is consistent with both lack of probabilistic sophistication and with probabilistically sophisticated preferences combined with non-expected utility risk preferences. This is reflected in the results: 9 out of the 49 subjects are probabilistically sophisticated (based on their choices in the standard Ellsberg problem and in One vs Two) and 7 of them exhibit CCE. Of the remaining 40, all of whom are not probabilistically sophisticated, 21 exhibit the CCE. Therefore, both interpretations have support in the data. Indeed, CCE is not significantly associated with probabilistic sophistication as reflected in Ellsberg standard problem and One vs Two.[18] However, since the great majority of subjects are *not* probabilistically sophisticated, 75% of the subjects who exhibit CCE violate probabilistic sophistication. This lends support to the interpretation that most occurrences of CCE result from subjects' attitude towards ambiguity about the relation between urns.

[16] Out of 56 subjects, the answers of 7 subjects to questions 9 and 10 were omitted. For two of them there was an error by the research assistants in inserting the conditional certainty equivalents in question 10 based on the responses to the previous question, and the rest had extremely low (0 or 1) or extremely high (99 or 100) switching points, which we thought did not make any economic sense. Since the first of the CCE questions involved a price list, while the rest of the questions involved only binary choices, we believe these choices resulted from a misunderstanding of the experimental protocol in this question and did not reflect on other questions.

[17] Note that 57% is a lower bound on the proportion of subjects exhibiting CCE, since the approximate conditional certainty equivalent used in question 10 is the largest integer x such that, for example: $(100, \{R_1B_2\}) \succsim (x, \{R_1B_2, B_1R_2\})$, and does not necessarily reflect indifference.

[18] CCE behavior is not associated with Ellsbergian ambiguity and with choices in One vs Two, measured separately.

3.2.4 Same versus Different

Though not the main focus of the experiment, we asked subjects to rank the bets $Same$ and $Diff$. Combining the latter ranking with the choices in One vs Two provides a way to test for nontransitivity (assuming monotonicity). For example, revealing that $Same \prec R_1 \prec Diff$ in questions 1-4 but choosing consistently with $Same \succ Diff$ in questions 5-6, contradicts transitive preference. Due to the complexity of the experiment, we felt that it was important to verify that the rankings obtained in One vs Two are not arbitrary. A second rationale for the inclusion of these questions is that the de Finetti exchangeable model and some generalizations described in the next section imply the (weak) preference for $Same$ over $Diff$. Therefore, the experimentally observed rankings reflect on these and other models examined below. A third rationale (discussed in subsection 3.2.5) for including this question is that it serves as a simple test whether the implementation of the RIS (using lotteries before balls are draws from the urns) eliminates potential hedging through the randomization used to incentivize subjects.

We find that slightly less than half of the subjects (27 subjects) were indifferent between the two bets, and another 7 subjects strictly preferred $Same$. The remaining 22 subjects (39%) strictly preferred $Diff$.

Furthermore, the association between the ranking of $Same$ and $Diff$ and measures of probabilistic sophistication (PS) is very tight.[19] Table 3.3 reports the association between probabilistic sophistication in the Ellsberg questions and the preference between $Same$ and $Diff$.

Out of the 15 subjects whose choices in the Ellsberg choice problem were consistent with probabilistic sophistication, 13 were indifferent between $Same$ and $Diff$, and out of 29 subjects that exhibited strict preference between $Same$ and $Diff$,[20] only 2 were consistent with probabilistic sophistication in the Ellsberg choice problem (p-value Fisher exact test = 0.00067 < 0.01). Though intuitively and also at a theoretical level, the ranking of

[19] The fact that indifference between $Same$ and $Diff$ and probabilistic sophistication in One vs Two are strongly associated is not surprising. The most frequent way to be consistent with probabilistic sophistication in the One vs Two problem, is not to exhibit strict preference between R_1 and each of $Same$ and $Diff$. But this implies that $Same$ and $Diff$ will most likely not be ranked by strict preference as well, and hence the tight association (p-value Fisher exact test = 0.0372). The reason for the qualification "most likely" is that we do not observe indifference, only the lack of strict preference.

[20] More than 75% of those who exhibited strict preference between $Same$ and $Diff$, strictly preferred the latter.

Same vs Different	**Ellsberg**		**Total**
	PS	**notPS**	
Same $\sim$ Diff	13	14	27
Same $\nsim$ Diff	2	27	29
Total	15	41	56
p-value Fisher exact test = 0.00067			

Table 3.3: Ellsbergian ambiguity and Same versus Different

$Same$ and $Diff$ does not seem connected to probabilistic sophistication (or its violation), empirically subjects who are not indifferent between $Same$ and $Diff$ are for the most part the same subjects who are not probabilistically sophisticated. This connection is one of the empirical facts confronting the models presented in the next section.

3.2.5 RIS and Hedging

As discussed earlier, the usage of RIS in ambiguity experiments is problematic, since it could provide the subject with an opportunity to hedge the uncertainty using the randomization device employed in the incentive system. For this reason we performed the randomization *before* subjects made their choices. It is an empirical question whether subjects conformed with this order when evaluating lotteries, and if they did not, whether they hedged.

Hedging might be revealed in different ways. For example, consider the two choices comprising One vs Two, and suppose, for simplicity, that the individual attaches probability 1/2 to payoffs being dependent on each of these questions. Then, a subject who is ambiguity averse and acts as if the randomization occurs after the balls are drawn, could choose $Same$ over R_1 in one question and $Diff$ over R_1 in another, and be left with the same state-independent payoff (in expected utility units) that she would obtain from betting on drawing red from Ellsberg's risky urn. Accordingly, if she is ambiguity averse in the Ellsberg problem, the preceding combined choices would be preferable to R_1, and thus we would observe $Two \succ One$. However, as discussed above, the latter is observed for only 4 subjects, and only one of these is ambiguity averse in the Ellsberg problem.

Similarly, in the questions that compare $Same$ to $Diff$ directly, hedging would imply the absence of a strict preference between $Same$ and $Diff$. This follows since if the subject had perceived the problem as if the randomization used for incentives occurred after the balls are drawn from the urns, then choosing the bet with a prize of \$101 in each of the two questions would leave the subject with a lottery that pays \$101 with probability 1/2 independently of the state, thus completely hedging the ambiguity. However, empirically more than half of the subjects exhibit a strict preference in these questions and, as indicated in Table 3.3, these include the large majority of subjects who are ambiguity averse (and thus might wish to hedge).

We conclude that we do not find empirical evidence that subjects used the RIS to hedge ambiguity.

4 Can This Behavior Be Modeled?

The results of the experiment argue against subjective expected utility theory and also against probabilistic sophistication. Expected utility maximization implies indifference in the Ellsberg problem, the rankings (2.3) in One vs Two, and is inconsistent with the CCE. Out of 49 subjects who completed all the questions, only 2 subjects were consistent with these predictions. If one considers probabilistically sophisticated preferences that need not be SEU, then 7 more subjects who exhibited CCE are added (a total of 18.4%). In the larger group of the 56 subjects whose choices in the Ellsberg and One vs Two problems were consistent with monotonicity and transitivity, only 10 (less than 18%) were probabilistically sophisticated. Consequently, we consider models of preferences that are not probabilistically sophisticated.

4.1 De Finetti-Based Models

Because it is the most commonly used model of preference for a setting with repeated experiments, we begin by taking the exchangeable Bayesian model as a benchmark and we consider how it and some generalizations perform.

The exchangeable Bayesian model is defined by expected utility maximization with predictive prior P on $S_1 \times S_2$ having the well known "conditionally i.i.d." form

$$P = \int_{\Delta(\{R,B\})} (\ell \otimes \ell)\, d\mu(\ell)\,. \tag{4.1}$$

Here ℓ is a generic probability law on S describing a single urn, and μ is a prior over these laws. Because it is probabilistically sophisticated, this model is inconsistent with One vs Two (2.2). Further, it implies indifference in (2.5) in contradiction to CCE, ambiguity neutrality in the Ellsberg choices, and also

$$Same \succeq R_1 \succeq Diff$$

There are only two subjects who made choices consistent with these predictions.

Thus we turn next to two generalizations that have been studied. Epstein and Seo (2010) study preference between uncertain prospects in a framework of repeated experiments. Like this paper, they offer behavioral critiques of the Bayesian exchangeable model. Their main contribution is to provide axiomatic characterizations of two alternative generalizations; see also Epstein and Seo (2013). A major difference from this paper is that they adopt an Anscombe-Aumann framework (where prizes are lotteries over money, for example), which is commonly used in axiomatic work because of the analytical power that it delivers, but which is also widely acknowledged as being less satisfactory than a Savage domain (where prizes can be amounts of money). One reason that is particularly pertinent is that typical applications of the Anscombe-Aumann domain, including by Epstein and Seo, assume that the decision-maker is an expected utility maximizer when ranking lotteries (risk), which limits the descriptive scope of the analysis. In contrast, we adopt a Savage domain and avoid comparably severe a priori restrictions on preferences. In part because of this difference in frameworks, the new behaviors that we introduce here, One$\succ$Two and CCE, do not have obvious counterparts in the previous work.

Multiplicity in priors (MP):[21] Replace the single prior μ in (4.1) by a set $\mathcal{M} \subset \Delta(\Delta(\{R, B\}))$ of prior beliefs about the composition of each urn. Each such prior μ induces a predictive prior over the two urns via (4.1). Denote by $\mathcal{P}_{MP}$ the set of predictive priors generated in this way, that is,

$$\mathcal{P}_{MP} = \left\{ \int_{\Delta(\{R,B\})} (\ell \otimes \ell) d\mu(\ell) : \mu \in \mathcal{M} \right\}. \tag{4.2}$$

Importantly, each P in $\mathcal{P}_{MP}$ is exchangeable.

The utility of any act f is computed in two stages. First, its certainty equivalent is computed using an expected utility calculation for each P in

[21] For axiomatic treatments of models of this sort see Epstein and Seo (2010, Model 1), Al Najjar and De Castro (2010), and Cerreia-Vioglio et al (2013).

$\mathcal{P}_{MP}$, and second these certainty equivalents are aggregated, where the aggregator can be very general. More precisely, for any act f, utility is defined by

$$U(f) = W\left(u^{-1} \circ \left(\int_{S_1 \times S_2} u(f)\, dP\right)_{P \in \mathcal{P}_{MP}}\right), \tag{4.3}$$

where u is a fixed strictly increasing vNM index, and W is an *aggregator*, restricted to be increasing so that

$$\begin{aligned} \int_{\Omega} u(f')\, dP &> \int_{\Omega} u(f)\, dP \text{ for all } P \in \mathcal{P}_{MP} \implies \\ U(f') &> U(f). \end{aligned}$$

For some purposes below we restrict attention to specifications that imply ambiguity aversion in Ellsberg's two-urn experiment. One could adopt for W any of the functional forms studied in the ambiguity literature and motivated by Ellsberg's experiments, including maxmin (Gilboa and Schmeidler, 1989), the smooth model (Nau, 2006; Klibanoff et al, 2005; Seo, 2009), and variational utility (Maccheroni et al, 2006). Each implies a different ranking of bets on a single urn, but, as shown below, they have in common the inability to model the behaviors One vs Two and CCE that are centered on the relation between urns. The reason is that though W is very general above, the set of predictive priors $\mathcal{P}_{MP}$ is limiting because it consists exclusively of exchangeable measures.[22]

In the first two special cases noted, utility functions have the respective forms

$$\begin{aligned} u \circ U(f) &= \min_{P \in \mathcal{P}_{MP}} \left(\int_{S_1 \times S_2} u(f)\, dP\right) \qquad (4.4) \\ &= \min_{\mu \in \mathcal{M}} \int_{\Delta(S)} \left[\int_{S_1 \times S_2} u(f)\, d(\ell \otimes \ell)\right] d\mu(\ell), \text{ and} \end{aligned}$$

$$\begin{aligned} U(f) &= \int_{\mathcal{P}_{MP}} \varphi \circ u^{-1} \left(\int_{S_1 \times S_2} u(f)\, dP\right) dm(P) \qquad (4.5) \\ &= \int_{\Delta(\Delta(S))} \varphi \circ u^{-1} \left(\int_{\Delta(S)} \left(\int_{S_1 \times S_2} u(f)\, d(\ell \otimes \ell)\right) d\mu\right) dm(\mu), \end{aligned}$$

[22] Conditions on $\mathcal{P}_{MP}$ and W that imply symmetry with respect to colors as in (2.1) are readily specified. Note, however, that the MP model contradicts One $\succ$ Two and CCE even without imposing such symmetry.

for some probability measure m on $\Delta(\Delta(S))$. The former implies Ellsbergian ambiguity aversion, as does the latter if φ is concave.

Multiplicity in likelihoods (ML): This model is based on Dempster (1967,8), where the building block "belief functions" are introduced, and on Epstein and Seo (2013), which provides a decision-theoretic axiomatization for the case of repeated experiments. We concentrate on a special case of the model that suffices to make the point.[23]

Utility is a special case of Gilboa and Schmeidler's (1989) maxmin utility with the set of predictive priors $\mathcal{P}_{ML}$ constructed as follows. Think of the perception that there is a subset of balls that is fixed across urns such that the probability of drawing red from this common component in either urn equals $p < \frac{1}{2}$, and similarly for blue. The remaining proportion $\kappa = 1 - 2p > 0$ of each urn is idiosyncratic and not understood at all. Thus the perception is that the proportion of each color in each urn lies in the interval $[p, 1-p]$. In particular, though the same interval applies to each urn, the compositions of these idiosyncratic components may differ across urns and therefore, the two compositions may differ. The perception is further that a draw from urn i reveals nothing about the idiosyncratic part of urn j, and any pattern of correlation across idiosyncratic components of urns is considered conceivable. More precisely, the set $\mathcal{P}_{ML}$ of predictive priors on $\{R_1, B_1\} \times \{R_2, B_2\}$ is given by

$$\begin{aligned}\mathcal{P}_{ML} = \; & p^2 \left[\Delta(R_1R_2) + \Delta(B_1B_2) + \Delta(R_1B_2) + \Delta(B_1R_2)\right] \\ & + p\kappa\Delta(R_1 \times \{R_2, B_2\}) + p\kappa\Delta(B_1 \times \{R_2, B_2\}) \\ & + p\kappa\Delta(\{R_1, B_1\} \times R_2) + p\kappa\Delta(\{R_1, B_1\} \times B_2) \\ & + \kappa^2\Delta((\{R_1, B_1\} \times \{R_2, B_2\}))\end{aligned} \tag{4.6}$$

Each term admits a simple interpretation along the lines of the above informal description. For example, the set $\Delta(R_1 \times \{R_2, B_2\})$ reflects ignorance about the composition of the idiosyncratic component of the second urn conditional on drawing red from the the subset of red balls in urn 1 that are common to both urns; this term has weight $p\kappa$ equal to the probability that the draw from the first urn is from the common component and the draw from the second urn is from its idiosyncratic component. Similarly, the final term

[23] See Appendix B for an outline of a different model that features multiple likelihoods and yet has different predictions regarding One vs Two.

expresses complete ignorance about the idiosyncratic components including about how they are connected across urns.

A number of features of this specification should be noted. First, the symmetry conditions (2.1) are implied. Second, $\mathcal{P}_{ML}$ contains both nonidentical product measures and also many nonproduct measures, reflecting that the individual is not confident either that the urns are identical or that they are uncorrelated. In the limiting case $\kappa = 0$, the two urns are purely risky and i.i.d. with common proportion $p = \frac{1}{2}$ for both red and blue. Also for $\kappa > 0$, there is a sense in which $\mathcal{P}_{ML}$ models "independence" of the two urns: The implied utility function satisfies the "product rule" $U_{ML}(f_1 \cdot f_2) = U_{ML}(f_1)\, U_{ML}(f_2)$ for any acts f_i over urn i, where $f_1 \cdot f_2$ is the pointwise product of f_1 and f_2 and hence is an act over $S_1 \times S_2$.[24]

Table 4.1 describes some of the relevant predictions of both the MP and ML models and juxtaposes them with the associated experimental findings.[25]

	One$\succ$Two	**CCE**	**Ellsbergian ambiguity aversion**	**Same$\succeq$Diff**
% subjects	41% (36 indifferent)	57% (out of 49)	66%	61%
MP	X (*)	X	$\surd$ (*)	$\surd$
ML	$\surd$	$\surd$	$\surd$	$\surd$ (indiff)
*refers to the Ellsberg (weakly) ambiguity averse special case				

Table 4.1: Marginal frequencies and predictions of the MP and ML models

In the table, an X ($\surd$) indicates that the model cannot accommodate (implies) the indicated property. For example, MP cannot rationalize CCE. The table focusses on the special case of the MP model that exhibits Ellsbergian ambiguity aversion (in the weak sense, including neutrality). Where this assumption is necessary to derive an implication we indicate it by an asterisk. With that restriction, MP cannot accommodate One$\succ$Two either. As a result, it accommodates only 2 more subjects beyond the Bayesian exchangeable model. In contrast, ML implies all three behaviors (given $\kappa > 0$).

[24] Epstein and Seo (2013) refer to such a utility function as an IID utility function. The reader is referred to their papers (2010, 2013) and also to Cuoso et al (1999) for elaboration.

[25] See Appendix A for proofs.

Thus, looking at each of these behaviors separately, ML seems to perform well and much better than MP. This relative performance is intuitive: concern with poorly understood heterogeneity is about robustness to likelihoods, not to priors.

However, on closer examination the ML model fails in accounting for the associations between the behaviors. It is important to note that for the ML model, the three properties being discussed are tied together: One is satisfied if and only if each of the others is satisfied.[26] However, this tight connection is not observed in the experiment: *Only 10 subjects satisfied One $\succ$ Two, CCE and Ellsbergian ambiguity aversion.*[27] That is, together with the 2 subjects whose choices are rationalized by the Bayesian exchangeable model, the ML model can rationalize the choices of only 12 subjects out of 49 (less than 25% of the subjects). This limitation of the model in part motivates consideration of a different kind of model in the next subsection.

The other limitation of ML that has been ignored to this point relates to the final column of Table 4.1. The ML model as we have defined it implies indifference between betting on $Same$ and $Diff$, which was exhibited by slightly less than 50% of subjects, while roughly another 12% strictly preferred $Same$ and the remaining 39% strictly preferred to bet on $Diff$. Though the choice between $Same$ and $Diff$ was not our motivating behavior, it is nevertheless related to One vs Two and the experimentally observed rankings provide another measuring stick for candidate models. For the ML model, two alternative generalizations can accommodate the observed heterogeneity in rankings.

As ML is defined above, there is certainty about the value of p. Thus the model generalizes the Bayesian i.i.d. model rather than de Finetti's exchangeable model. More generally, the individual might be uncertain about p and have beliefs represented by μ^*, a measure over $[0, \frac{1}{2}]$, the set of possible values of p. For simplicity let μ^* have finite support, say $\{p_1, ..., p_N\}$. Her set $\mathcal{P}^*_{ML}$ of predictive priors would then be the mixture

$$\mathcal{P}^*_{ML} = \Sigma_{i=1}^N \mu^* (p_i) \, \mathcal{P}^i_{ML},$$

where $\mathcal{P}^i_{ML}$ corresponds to p_i. This is an exchangeable version of ML that is closer to the general model axiomatized in Epstein and Seo (2013). It is

[26] This can be seen from the proofs in Appendix A.

[27] As noted above, the MP performs even worse as there are only 2 subjects who are Ellsbergian ambiguity averse, indifferent between One and Two (or exhibit $Same \succeq R_1 \succeq Diff$) and do not exhibit CCE.

readily verified that with this extension ML implies a weak preference for $Same$, where the preference is strict except for knife-edge cases.[28] This can be understood by noting the parallel with the Bayesian exchangeable model, wherein the i.i.d. special case implies indifference between $Same$ and $Diff$, while the exchangeable extension implies that experiments are viewed as being positively correlated. The preceding are direct analogues for the ML model.[29]

Out of 56 subjects, 22 strictly preferred $Diff$ to $Same$. A Bayesian exhibits such a ranking if, for example, her predictive prior is P^{compl}, where

$$P^{compl}\left(R_1 B_2\right) = P^{compl}\left(B_1 R_2\right) = \tfrac{1}{2},$$

which can be understood in terms such as "there are only so many red balls to go around", or "regression to the ex-ante mean." This prior is symmetric across urns (it satisfies $P\left(R_1 B_2\right) = P\left(B_1 R_2\right)$), though it cannot be expressed in the form (4.1). The reason is that de Finetti's celebrated theorem characterizing exchangeability assumes infinitely many experiments.[30] Thus use of (4.1) presumes that the individual behaves as if she sees the two urns as part of an infinite sequence of symmetric urns, which appears contradicted by the experimental findings. However, the strict preference for $Diff$ can be accommodated by extending the ML model to permit the individual to attach positive probability (no matter how small) to complementarity between urns as above. For example, consider the set $\mathcal{P}_{ML}^{**}$ of predictive priors given by

$$\mathcal{P}_{ML}^{**} = (1-\epsilon)\,\mathcal{P}_{ML} + \epsilon\{P^{compl}\},$$

where $\epsilon > 0$ is a fixed parameter. Then $Diff \succ Same$ and, for sufficiently small ϵ, the implications of ML for One vs Two, CCE and Ellsberg aversion remain intact (by continuity).

Conclude that the ML model, including both the original formulation and also the two extensions just described, can, if one allows preference heterogeneity, rationalize all rankings of $Same$ vs $Diff$ without affecting

[28] Moreover, Ellsbergian aversion, One $\succ$ Two and CCE are still implied.

[29] Similarly, the MP model implies that $Same$ is weakly preferred, because it can be understood as aggregating a number of different exchangeable Bayesian selves each of whom prefers $Same$.

[30] He also provides a representation result when there are only finitely many experiments. However, that result has not yet been extended to nonprobabilistically sophisticated preferences.

the explanatory power with regard to the other three behaviors in Table 4.1. Thus the inability to model the imperfect connections between the latter properties that one observes in the experimental results remains the central challenge for this model.

Remark 4.1 *The ML model and its extensions assume more than just maxmin plus symmetry, and it is to be expected that a broader framework, given also preference heterogeneity, would provide added flexibility. For example, consider the maxmin model with set of predictive priors consisting of all measures P of the form*

$$P(R_1B_2) = P(B_1R_2) = \tfrac{1}{4} + \delta \text{ and } P(R_1R_2) = P(B_1B_2) = \tfrac{1}{4} - \delta,$$

where δ varies over $[-a, a]$ and a is a fixed parameter, $0 < a < \frac{1}{4}$. Then $P(R_i) = P(B_i) = \frac{1}{2}$ for each urn $i = 1, 2$, which implies indifference between betting on R_1 and on the risky Ellsberg urn. But the measures P disagree about the joint distributions across the two urns, which leads to the preference One $\succ$ Two. Thus maxmin can rationalize Ellsbergian ambiguity neutrality simultaneously with a strict preference for One over Two. The example implies also the indifference $f_1 \sim x$ in (2.5), thus violating CCE and predicting a pattern of behavior exhibited by only two subjects.[31] See Appendix B for more on what can be done within the maxmin framework to rationalize observed rankings if one allows preference heterogeneity.

4.2 A Source-Based Model

This model acknowledges that there are three sources of uncertainty, or issues, in the experiment. Two of them are familiar: *risk*–when the composition of the urn is known, and *bias*–when the composition of the urn is unknown; they have been studied extensively both theoretically and empirically since Ellsberg proposed his thought experiments. The contrast between them has become synonymous with "Knightian uncertainty" and the distinction between risk and ambiguity. The new issue, which to the best of our knowledge has not been studied before, is the *heterogeneity and relation* between urns. Note that the latter issue is excluded a priori in an Ellsberg-style

[31] A similar example is given in Epstein and Seo (2010, Example 4.3), where it is shown that it violates an axiom, Orthogonal Independence, that is central to their main representation result (Theorem 5.2).

risk-versus-ambiguity experiment because the composition of one of the urns is known and hence the draw from this urn is naturally taken to be completely independent of the draw from the ambiguous urn. However, when the decision-maker faces two ambiguous urns, her choice between bets on the draws from both urns may depend on and reveal her concern with how they may differ and be related to one another.

We outline a streamlined model that relates to our specific experiment; readers will see that the model can be generalized in a number of directions. First, we elaborate on the three issues; Figure 4.1 provides a diagrammatic representation. Notationally, the composition of a single urn is described by a probability vector of the form $(p, 1-p)$, where p denotes the proportion of red, and the joint composition of the two urns is described by a probability vector of the form $(p_{RB}, p_{BR}, p_{RR}, p_{BB})$. If urns 1 and 2 are described by $(p, 1-p)$ and $(q, 1-q)$ respectively, then $(p, 1-p) \otimes (q, 1-q)$ denotes the joint distribution given by the product measure,

$$(p, 1-p) \otimes (q, 1-q) \equiv (p(1-q), q\,(1-p)\,, pq, (1-p)\,(1-q))\,.$$

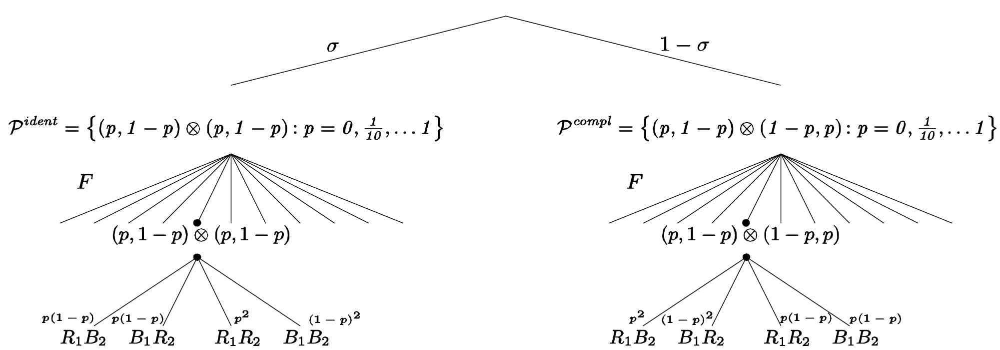

Figure 4.1: Hierarchical beliefs on the two ambiguous urns

Issue 1 Uncertainty about the relation between urns takes the form of two alternative hypotheses. One possibility entertained by the decision-maker is that the urns are i.i.d. according to $(p, 1-p)$ for some (unspecified) $p \in \left\{\frac{k}{10} : k = 0, ..., 10\right\}$, that is, the two urns can be described by a measure in the set $\mathcal{P}^{ident} = \left\{(p, 1-p) \otimes (p, 1-p) : p = 0, \frac{1}{10}, ..., 1\right\}$. The alternative is that the urns are "complementary" in the sense that their joint distribution lies in the set $\mathcal{P}^{compl} = \left\{(p, 1-p) \otimes (1-p, p) : p = 0, \frac{1}{10}, ..., 1\right\}$. This hypothesis is justified, for example, by the following perception of how the urns are constructed: there are 20 balls that are either red or blue–ten are drawn without replacement to fill urn 1 and the remaining ten are put into urn 2. The two hypotheses are assigned subjective probabilities σ and $1 - \sigma$ respectively.

Issue 2 Conditioning on either of the above hypotheses, the composition, or bias, of each urn is uncertain. Thus the decision-maker forms (probabilistic) beliefs about p. Conditional on each hypothesis, she uses the cdf F over possible values of p.

Issue 3 After conditioning on both the relation between urns and on the bias (through p), there remains uncertainty about the colors of the two drawn balls. However, resolution of Issues 1 and 2 implies a unique probability distribution over $\{R_1B_2, B_1R_2, R_1R_2, B_1B_2\}$. Thus the last issue concerns risk.

Each bet (or act) f associates a (dollar) payoff to each terminal node, depending on the colors of the two balls drawn. Thus it induces a 3-stage lottery, denoted D_f. A utility function Ψ over such lotteries is the final component defining preference over bets; more specifically, the utility of f is given by

$$U(f) = \Psi(D_f).$$

The utility of a bet on Ellsberg's risky urn can also be computed using Ψ. Such a bet involves only risk. Therefore, evaluate it via the induced (single-stage) lottery that the bet induces, using the restriction of Ψ to single-stage lotteries that are resolved completely at the third stage. For example, the bet on drawing red from the risky urn has utility $\Psi(Q_1)$, where $Q_1 \in \Delta(\Delta(\Delta(\mathbb{R})))$ is the three-stage lottery given by $Q_3 = \left(100, \frac{1}{2}; 0, \frac{1}{2}\right)$, $Q_2(Q_3) = 1$ and $Q_1(Q_2) = 1$.

Importantly, by not insisting that multistage lotteries be reduced according to the usual probability calculus, the model permits a (partial) disentangling of attitudes towards the three issues. We assume a functional form for Ψ along the lines of Kreps and Porteus (1978), with an expected utility function at each stage, but with different utility indices for different issues, and with compound lotteries evaluated recursively.[32]

To define utility precisely, let u_1, u_2 and u_3 be strictly increasing vNM indices that will apply to the three sources respectively. For any act f over $S_1 \times S_2$, its utility (in certainty equivalent units) is computed recursively by:

$$U(f) = u_1^{-1}\left(\sigma u_1(V_2(f)) + (1-\sigma)u_1(W_2(f))\right),$$

$$V_2(f) = u_2^{-1}\left(\int_p u_2(V_3(f;p))\,dF\right),\ W_2(f) = u_2^{-1}\left(\int_p u_2(W_3(f;p))\,dF\right),$$

$$\begin{aligned} V_3(f;p) &= u_3^{-1}\left(\int_{S_1\times S_2} u_3(f)\,d\left[(p,1-p)\otimes(p,1-p)\right]\right), \text{ and} \\ W_3(f;p) &= u_3^{-1}\left(\int_{S_1\times S_2} u_3(f)\,d\left[(p,1-p)\otimes(1-p,p)\right]\right). \end{aligned}$$

Recall the behavioral expression of symmetry in (2.1). Symmetry in urns is built into the model, and symmetry in colors is implied if we assume, as we do, that F is suitably symmetric (the compositions $(p, 1-p)$ and $(1-p, p)$ are equally likely). For the other behaviors of interest, note that several relevant bets depend only on subsets of issues. For example, Ellsbergian ambiguity aversion in this framework requires that the decision-maker prefer a multi-stage lottery that resolves at the last stage (risk) over a lottery that resolves over the second (uncertainty about bias) and last (risk) stages. This is true if and only if u_2 is more concave than u_3. In comparing One vs Two, the bet R_1 is subject to uncertainty about the bias; its payoff does not depend on the first issue because R_1 is a bet on only one urn, but the payoffs to both $Same$ and $Diff$ depend on the relation between urns. This suggests that the ranking of One vs Two depends on both the relative curvatures of u_1 and u_2 and on the magnitude of σ. A characterization in complete generality is not available, nor for CCE. However, a special case that we outline next suffices to demonstrate the flexibility of this model in separating behaviors. (See Appendix C for supporting details.)

Specialize the model first by assuming that the cdf F is given by

$$F = \begin{cases} 1/3 & 0 \le p < \frac{1}{2} \\ 2/3 & \frac{1}{2} \le p < 1 \\ 1 & p = 1 \end{cases} \tag{4.7}$$

[32] An alternative, following Segal (1987,1990), would be to use non-expected utility functions at each stage.

Ellsberg	**One vs Two**	**Total**	**CCE**	
			Yes	**No**
Averse	One $\succ$ Two	15 $0 < \alpha_2 < \alpha_1$	10 $0 < \alpha_2 < \alpha_1 < 2^{\dagger}$	5 $0 < \alpha_2 \ll \alpha_1^{\dagger}$
Averse	One $\sim$ Two	9 $0 < \alpha_1 = \alpha_2$	8 $0 < \alpha_1 = \alpha_2 < 2^{\dagger}$	1 $\nexists \alpha_1, \alpha_2$
Neutral	One $\not\sim$ Two	5 $0 = \alpha_2 \neq \alpha_1$	1 $0 = \alpha_2 < \alpha_1 < 2^{\dagger}$	4 $0 = \alpha_2 \ll \alpha_1^{\dagger}$
Neutral	One $\sim$ Two	9 $\alpha_1 = \alpha_2 = 0$	7 $\nexists \alpha_1, \alpha_2$	2 $\alpha_1 = \alpha_2 = 0$

Table 4.2: Empirical frequencies of the main behavioral patterns and corresponding sufficient parametric restrictions in the source-based model

that is, each urn contains either 10 red balls, 10 blue balls, or 5 of each color, and each possibility is equally likely. Second, let $\sigma = \frac{1}{2}$, which can be shown to imply (indeed, characterize) indifference between *Same* and $Diff$;[33] we relax this restriction below. With regard to utility indices, let u_3 be linear (risk neutrality) and adopt the normalization $u_3(1) = 1$, $u_3(0) = 0$.[34] Finally, let u_1 and u_2 be isoelastic:

$$u_i(x) = \begin{cases} \frac{x^{1-\alpha_i}}{1-\alpha_i} & \text{if } \alpha_i \neq 1 \\ \log x & \alpha_i = 1 \end{cases}, \quad i = 1, 2,$$

where α_i is positive and measures the concavity of u_i (or aversion to the uncertainty in issue i). For the calculations that follow, we assume also that

$$\alpha_2 < 1.$$

Under these assumptions, elementary calculations (Appendix C) show that One $\succ$ Two if $\alpha_1 > \alpha_2$. We use the special case more fully in Table 4.2 to describe how the model can relax the tight connections imposed by de Finetti-based models, and thus can accommodate several behavioral patterns found in the data.

[33] Similarly, *Same* is strictly preferred if and only if $\sigma > \frac{1}{2}$.

[34] Under risk aversion we suspect that some of the findings below generalize if concavity of each of u_1 and u_2 is measured relative to the concavity of u_3; but this remains a conjecture.

The parameter restrictions marked with † are based in part on Figure C.1 in Appendix C. Accordingly, $\alpha_2 \ll \alpha_1$ means that the pair (α_1, α_2) lies in the lightly shaded region, which is the region where CCE is violated. The reader will see from Figure C.1 that some parameter restrictions in the Table (such as $\alpha_1 < 2$) can be relaxed; we opted for greater simplicity instead of generality.

The results reported in Table 4.2 are based on the assumption that $\sigma = \frac{1}{2}$, which, as noted, implies indifference between *Same* and $Diff$. But about one half of subjects violate this indifference. To accommodate also this dimension of behavior, note that by increasing σ slightly above $\frac{1}{2}$, the model can rationalize the strict preference for *Same*, while at the same time (by continuity) not changing any of the other strict rankings indicated, such as Ellsbergian aversion, One $\succ$ Two, or CCE.[35] The strict preference for $Diff$ can be accommodated analogously by taking σ slightly smaller than $\frac{1}{2}$. In this way, the model can rationalize the most commonly observed behavioral patterns found in the data.

The two rankings[36]

$$Same \succ R_1 \succ Diff, \text{ or } Diff \succ R_1 \succ Same.$$

cannot be rationalized by small perturbations of σ about $\frac{1}{2}$. However, the model can rationalize these choice patterns if we take σ sufficiently different from $\frac{1}{2}$. For example, if σ is sufficiently close to 0, then the individual is extremely confident that the urns are complementary and this leads to the ranking $Diff \succ R_1 \succ Same$; moreover, this ranking can prevail even though she may be ambiguity averse in the Ellsbergian sense and regardless of the relative magnitudes of α_1 and α_2. In the same way, the ranking $Same \succ R_1 \succ Diff$ can be rationalized if σ is sufficiently close to 1.

Another behavioral pattern that is not accommodated in Table 4.2 is the combination of Ellsberg neutrality, One$\sim$Two and CCE. As noted earlier, these rankings may be rationalized by probabilistically sophisticated preferences exhibiting the Allais certainty effect.

[35] We used a similar perturbation for the de Finetti-based models. Note also that ambiguity neutrality is unaffected by the value of σ, and that One $\sim$ Two means that R_1 is indifferent to both *Same* and $Diff$ and therefore that it can be exhibited only if $Same \sim Diff$, and thus $\sigma = \frac{1}{2}$.

[36] Two subjects chose according to the former while 7 chose according to the latter.

5 Concluding Remarks

The literature stimulated by Ellsberg, particularly the experimental literature, has focussed on the two-fold distinction between risk and ambiguity, or 'Knightian uncertainty.' In Ellsberg's two-urn experiment, the latter is embodied in the uncertain bias (or composition) of the unknown urn. In a setting with repeated experiments, or multiple ambiguous urns, we have introduced a second source of ambiguity–correlation between experiments. Thus we have studied at a behavioral level the three-fold distinction between risk, bias and correlation.

There are other thought and laboratory experiments in the literature using multiple urns or sources of ambiguity. The most prominent, of course, is Ellsberg's classic two-urn experiment which we have already contrasted with ours. Eliaz and Ortoleva (2012) and Eichberger et al. (2012) conduct experiments where there are multiple dimensions of ambiguity (for example, an ambiguous probability of winning and an ambiguous winning prize). They also investigate the association between (ambiguity averse) behaviors in different dimensions, but they do not have counterparts of our behavioral hypotheses. Multiple urns are used in experiments exploring learning and dynamic consistency. These issues are not involved in our study because we consider only ex ante choice.

An experiment that has not been conducted but that is potentially useful for evaluating the arguments or interpretations proposed in the current paper is worth mentioning. We have interpreted the preference One $\succ$ Two as reflecting (an aversion to) ambiguity about how urns are related. An alternative explanation is that subjects may simply dislike bets that depend on two draws relative to bets that depend only on a single draw. To explore this hypothesis, one might consider behavior in the extreme case where the two draws are made (with replacement) from a single ambiguous urn. The finding that $R_1 \succ Same$ where the two draws come from the same urn would challenge our interpretation of the results in the current study.

A Proofs for Table 4.1

This appendix provides proofs for the assertions in Table 4.1. Here and in the next appendix, it is without loss of generality to adopt the normalization

$$u(0) = 0 \text{ and } u(100) = 1.$$

MP model:

(MP.i) MP violates One $\succ$ Two if it exhibits weak Ellsberg ambiguity aversion: Because $p^2 + (1-p)^2 \geq \frac{1}{2}$ for all p in $[0,1]$, infer that

$$P(Same) = \int \left[(\ell(R))^2 + (\ell(B))^2 \right] d\mu(\ell) \geq \tfrac{1}{2}$$

for every exchangeable P. But $\mathcal{P}_{ML}$ contains only exchangeable measures and W is increasing. Therefore, $Same$ is weakly preferred to betting on red in a 50-50 urn. If weakly Ellsberg averse, then $Same \succeq R_1$, contrary to One $\succ$ Two.

(MP.ii) MP violates CCE: Because every predictive prior P is exchangeable and W is increasing,

$$\begin{array}{ccc}
\begin{bmatrix} 1 & R_1B_2 \\ 0 & B_1R_2 \\ 0 & R_1R_2 \\ 0 & B_1B_2 \end{bmatrix} & \sim & \begin{bmatrix} x & R_1B_2 \\ x & B_1R_2 \\ 0 & R_1R_2 \\ 0 & B_1B_2 \end{bmatrix} \Longleftrightarrow u(x) = \dfrac{1}{2} \Longrightarrow \\
\displaystyle\int_{S_1 \times S_2} u(f_1)\, dP & = & \displaystyle\int_{S_1 \times S_2} u(g_1)\, dP \quad \text{for every } P \Longrightarrow \\
\begin{bmatrix} 1 & R_1B_2 \\ 0 & B_1R_2 \\ x & R_1R_2 \\ x & B_1B_2 \end{bmatrix} & \sim & \begin{bmatrix} x & R_1B_2 \\ x & B_1R_2 \\ x & R_1R_2 \\ x & B_1B_2 \end{bmatrix}.
\end{array}$$

(MP.iii) MP implies $Same \succeq Diff$: For any $P = \int_{\Delta(S)} (\ell \otimes \ell) d\mu(\ell)$,

$$\begin{aligned}
P(Same) &= P(R_1R_2) + P(B_1B_2) = \int \left[(\ell(R))^2 + (\ell(B))^2 \right] d\mu(\ell) \\
&\geq 2 \int \ell(R)\, \ell(B)\, d\mu(\ell) = P(R_1B_2) + P(B_1R_2) = P(Diff).
\end{aligned}$$

Since this inequality is satisfied by every P in $\mathcal{P}_{ML}$, and since W is increasing, $U(Same) \geq U(Diff)$.

ML model:

(ML.i) ML implies One $\succ$ Two: Proof is by straightforward calculation of minima over $\mathcal{P}_{ML}$ given by (4.6). The latter is a mixture of sets, and it is important to note that minimization can be performed over each component set separately. Thus, for example, when computing the utility of $Same$, one contribution to this utility is through the minimum probability of $Same$ as one varies over $\Delta(R_1 \times \{R_2, B_2\})$, which minimum equals 0. In contrast, when evaluating R_1, the minimum probability over $\Delta(R_1 \times \{R_2, B_2\})$ equals 1. This "explains" why the bet on one urn is strictly preferred. Similarly when comparing R_1 with $Diff$. More specifically, calculate that

$$U_{ML}(\{R_1B_2, B_1R_2\}) = U_{ML}(\{R_1R_2, B_1B_2\}) = 2p^2 < 2p^2 + p\kappa = U_{ML}(R_1).$$

(ML.ii) ML implies CCE: Compute that $U_{ML}(f_0) = U_{ML}(g_0) \implies u(x) = 1/2 > U_{ML}(f_1) = p$.

(ML.iii) ML implies that $Same \sim Diff$: As noted above, both have utility equal to $2p^2$.

B A Different Multiple-Likelihood Model

We describe an alternative model using multiple likelihoods, (see Walley and Fine (1982), and Epstein and Seo (2010)), and we show that while it accommodates CCE, it cannot rationalize One $\succ$ Two. Therefore, the behaviors we study discriminate also between two models that both feature multiple likelihoods.

Utility is a special case of Gilboa and Schmeidler's (1989) maxmin utility having the set of predictive priors $\mathcal{P}_{prod}$ constructed as follows. Fix $0 < p < \frac{1}{2}$ and, as in the ML model, suppose that for each urn the probability of red is thought to lie in the interval $[p, 1-p]$; within the interval there is ignorance. The difference from ML is that the individual entertains only probability laws that are (nonidentical) products of measures on each urn. More precisely, $\mathcal{P}_{prod}$ consists of all product measures on $S_1 \times S_2$ of the form $\ell_1 \otimes \ell_2$ such that each ℓ_i is a measure on S such that $p \leq \ell_i(R) \leq 1-p$. If we abuse notation and denote this set of measures by $[p, 1-p]$, then one might write

$$\mathcal{P}_{prod} = [p, 1-p] \otimes [p, 1-p].$$

The set $\mathcal{P}_{prod}$ contains many nonidentical product measures, which suggests the capacity to capture the possibility that the urns differ from one another. However, as is readily verified,

$$\mathcal{P}_{prod} \subset \mathcal{P}_{ML},$$

which is interpretable as the present model implying less ambiguity (aversion) about how experiments differ and/or are related than the ML model. This difference is responsible for their differing predictions described next.

The two multiple likelihood models have in common that they predict Ellsberg ambiguity aversion and CCE, but *they differ in their predictions regarding One vs Two*. For the present model, we have

$$Same \sim Diff \succ R_1.$$

Following the discussion of One vs Two in Section 2, we interpret the strict inferiority of R_1 as indicating a greater aversion to ambiguity about the bias of any single urn than to ambiguity about differences between urns.

Here is a sketch of the elementary proof:

CCE is satisfied: Compute that $f_0 \sim g_0$ iff

$$u\left(x\right) = \tfrac{1}{2}\frac{p}{1-p} < \tfrac{1}{2}.$$

It remains therefore to show that

$$U_{prod}\left(\left[\begin{array}{cc} 1 & R_1B_2 \\ 0 & B_1R_2 \\ x & R_1R_2 \\ x & B_1B_2 \end{array}\right]\right) < u\left(x\right) = \tfrac{1}{2}\frac{p}{1-p}.$$

Compute

$$\begin{aligned} & \min_{p_1,p_2}\left[p_1\left(1-p_2\right) + u\left(x\right)\left(p_1p_2 + \left(1-p_1\right)\left(1-p_2\right)\right)\right] \\ = \; & \min_{p_1,p_2}\left[p_1\left(1-p_2\right) + u\left(x\right)\left(2p_1p_2 - p_1 - p_2\right) + u\left(x\right)\right]. \end{aligned}$$

Therefore, the behavior is rationalized iff

$$\min_{p_1,p_2}\left[p_1\left(1-p_2\right) + u\left(x\right)\left(2p_1p_2 - p_1 - p_2\right)\right] < 0.$$

But LHS is no greater than (taking $p_1 = 1 - p_2 = p$)

$$\begin{aligned} & \left[p^2 + u\left(x\right)\left(2p(1-p) - p - (1-p)\right)\right] \\ = \quad & p^2 - u\left(x\right)\left(2p^2 - 2p + 1\right) \\ = \quad & p^2 - \tfrac{1}{2}\frac{p}{1-p}(2p^2 - 2p + 1) \\ = \quad & p\left[p - \tfrac{1}{2}\frac{1}{1-p}(2p^2 - 2p + 1)\right] \; < 0 \\ \Longleftrightarrow \quad & -\left(2p-1\right)^2 < 0. \end{aligned}$$

$Same \sim Diff \succ R_1$:

$$\begin{aligned} U_{prod}\left(\{R_1R_2, B_1B_2\}\right) &= U_{prod}\left(\{B_1R_2, R_1B_2\}\right) = \\ 2p(1-p) &> p = U\left(R_1\right). \end{aligned}$$

No subjects exhibit Ellsbergian aversion, CCE and the strict preference for betting on both urns rather than on one. However, besides illustrating that the connection between "multiple likelihoods" and the ranking One $\succ$Two is complex, the model just described can play a role also in rationalizing the more common behavior (exhibited by 8 subjects) consisting of Ellsbergian ambiguity aversion, CCE and One $\sim$ Two.

To see this, define a new set of predictive priors by

$$\mathcal{P} = \tfrac{1}{2}\mathcal{P}_{ML} + \tfrac{1}{2}\mathcal{P}_{prod},$$

which yields the maxmin utility function U given by

$$U = \tfrac{1}{2}U_{ML} + \tfrac{1}{2}U_{prod}.$$

An interpretation is that the individual is certain that the proportion of red balls in each urn lies in $[p, 1-p]$, (which interval is common to both U_{ML} and U_{prod}), but is uncertain about which utility or set of priors describes the relation between urns, and she attaches equal probability to each.[37] Then U is ambiguity averse in the Ellsbergian sense (because both U_{ML} and U_{prod} are), satisfies CCE (as can be verified), and U implies indifference between betting on one urn or on two; more precisely,

$$Same \sim Diff \sim R_1.$$

(The "explanation" is that the opposite rankings implied by U_{ML} and U_{prod} perfectly offset one another.[38])

Remark B.1 *Like the basic ML model, both U_{prod} and U assume certainty about the correct value of p, and as a result generalizes the Bayesian i.i.d. model. Extensions analogous to those described at the end of Section 4.1 can be formulated also here for U_{prod} and/or U so as to accommodate any desired ranking of Same versus Diff without affecting the other behaviors described above.*

[37] This is a special case of the axiomatic model in Epstein and Seo (2010, Thm. 5.2) which is formulated for the case of infinitely many urns or experiments.

[38] Compute that $U_{ML}(R_1) = U_{prod}(R_1) = p$, $U_{ML}(Same) = U_{ML}(Diff) = 2p(1-p) > p$, and $U_{prod}(Same) = U_{prod}(Diff) = 2p^2$.

C Utilities in the Source-Based Model

We provide some details here supporting the discussion of the source-based model (Section 4.2), particularly for Table 4.2. Assume u_3 linear throughout, and for later calculations consider the special case, including concave isoelastic utility indices.

The utility of the bet R_3 on drawing red from the risky Ellsberg urn is given by

$$U(R_3) = u_3^{-1}\left(\tfrac{1}{2}\right) = \tfrac{1}{2},$$

and the utility of the bet R_1 on drawing red from a single ambiguous urn is

$$U(R_1) = u_2^{-1}\left(\int_p u_2(p)\, dF\right).$$

Thus R_3 is preferred if u_2 is concave.

For the bets *Same* and *Diff*, compute that $u_1 \circ U(Same)$ and $u_1 \circ U(Diff)$ are given respectively by

$$\sigma u_1 \circ u_2^{-1}\left[\int_p u_2(1-2p(1-p))\, dF\right] + (1-\sigma) u_1 \circ u_2^{-1}\left[\int_p u_2(2p(1-p))\, dF\right],$$
$$\sigma u_1 \circ u_2^{-1}\left[\int_p u_2(2p(1-p))\, dF\right] + (1-\sigma) u_1 \circ u_2^{-1}\left[\int_p u_2(1-2p(1-p))\, dF\right].$$

Since $1-2p(1-p) \geq 2p(1-p)$ for all p, it follows that *Same* is preferred if and only if $\sigma > \frac{1}{2}$ and that they are indifferent if $\sigma = \frac{1}{2}$.

Focus now on the special case, including the specification (4.7) for F, isoelastic utilities, $\alpha_2 < 1$ and $\sigma = \frac{1}{2}$. Then One $\succ$ Two is implied if $\alpha_1 > \alpha_2$. For example,

$$\begin{aligned} u_1 \circ U(Same) &= \tfrac{1}{2} u_1 \circ u_2^{-1}\left[\tfrac{2}{3} u_2(1) + \tfrac{1}{3} u_2\left(\tfrac{1}{2}\right)\right] + \tfrac{1}{2} u_1 \circ u_2^{-1}\left[\tfrac{2}{3} u_2(0) + \tfrac{1}{3} u_2\left(\tfrac{1}{2}\right)\right] \\ &\leq u_1 \circ u_2^{-1}\left[\tfrac{1}{2}\left(\tfrac{2}{3} u_2(1) + \tfrac{1}{3} u_2\left(\tfrac{1}{2}\right)\right) + \tfrac{1}{2}\left(\tfrac{2}{3} u_2(0) + \tfrac{1}{3} u_2\left(\tfrac{1}{2}\right)\right)\right] \\ &= u_1 \circ u_2^{-1}\left[\tfrac{1}{3} u_2(0) + \tfrac{1}{3} u_2\left(\tfrac{1}{2}\right) + \tfrac{1}{3} u_2(1)\right] = u_1 \circ U(R_1). \end{aligned}$$

Finally, examine CCE ((2.4) and (2.5)). Compute that $u_1 \circ U(f_0) =$

$$\tfrac{1}{2} u_1 \circ u_2^{-1}\left[\tfrac{2}{3} u_2(0) + \tfrac{1}{3} u_2\left(\tfrac{1}{4}\right)\right] + \tfrac{1}{2} u_1 \circ u_2^{-1}\left[\tfrac{1}{3} u_2(0) + \tfrac{1}{3} u_2\left(\tfrac{1}{4}\right) + \tfrac{1}{3} u_2(1)\right],$$

and that $u_1 \circ U(g_0) =$

$$\tfrac{1}{2} u_1 \circ u_2^{-1}\left[\tfrac{2}{3} u_2(0) + \tfrac{1}{3} u_2\left(\tfrac{x}{2}\right)\right] + \tfrac{1}{2} u_1 \circ u_2^{-1}\left[\tfrac{2}{3} u_2(x) + \tfrac{1}{3} u_2\left(\tfrac{x}{2}\right)\right],$$

which implies that

$$
\begin{aligned}
f_0 \sim\ & g_0 \Longleftrightarrow \\
& u_1 \circ u_2^{-1}\left[\tfrac{1}{3}u_2\left(\tfrac{1}{4}\right)\right] + u_1 \circ u_2^{-1}\left[\tfrac{1}{3} + \tfrac{1}{3}u_2\left(\tfrac{1}{4}\right)\right] = \\
& u_1 \circ u_2^{-1}\left[\tfrac{1}{3}u_2\left(\tfrac{x}{2}\right)\right] + u_1 \circ u_2^{-1}\left[\tfrac{2}{3}u_2\left(x\right) + \tfrac{1}{3}u_2\left(\tfrac{x}{2}\right)\right],
\end{aligned}
$$

and hence

$$
x = \left[\frac{\left(\frac{1}{3}\left(\frac{1}{4}\right)^{1-\alpha_2}\right)^{\frac{1-\alpha_1}{1-\alpha_2}} + \left(\frac{1}{3}\left(\frac{1}{4}\right)^{1-\alpha_2} + \frac{1}{3}\right)^{\frac{1-\alpha_1}{1-\alpha_2}}}{\left(\frac{1}{3}\left(\frac{1}{2}\right)^{1-\alpha_2}\right)^{\frac{1-\alpha_1}{1-\alpha_2}} + \left(\frac{2}{3} + \frac{1}{3}\left(\frac{1}{2}\right)^{1-\alpha_2}\right)^{\frac{1-\alpha_1}{1-\alpha_2}}}\right]^{\frac{1}{1-\alpha_1}}.
$$

In addition,

$$
\begin{aligned}
u_1 \circ U\left(f_1\right) &= \tfrac{1}{2}u_1 \circ u_2^{-1}\left[\tfrac{2}{3}u_2\left(x\right) + \tfrac{1}{3}u_2\left(\tfrac{1}{4} + \tfrac{x}{2}\right)\right] \\
&\quad + \tfrac{1}{2}u_1 \circ u_2^{-1}\left[\tfrac{1}{3}u_2\left(0\right) + \tfrac{1}{3}u_2\left(1\right) + \tfrac{1}{3}u_2\left(\tfrac{1}{4} + \tfrac{x}{2}\right)\right], \text{ and} \\
u_1 U\left(g_1\right) &= u_1 \circ u_2^{-1}\left[u_2\left(x\right)\right] = u_1\left(x\right), \text{ or } U\left(g_1\right) = x.
\end{aligned}
$$

We do not have an analytical characterization of the pairs (α_1, α_2) for which $U(f_1) < x$ as required by CCE. However, by numerical means we derive Figure C.1 which provides a clear picture of the regions in parameter space where CCE is and is not satisfied.

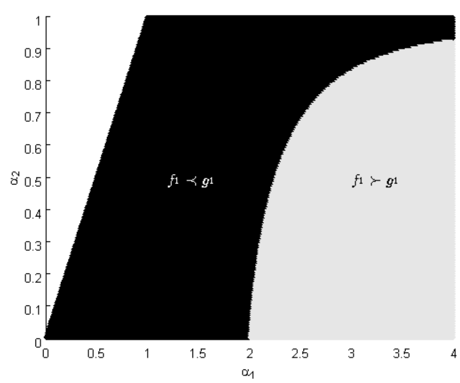

Figure C.1: Ranking of f_1 and g_1 as a function of the curvatures of isoelastic utility indices ($\alpha_1 \geq \alpha_2$)

D Weaker Inclusion Criteria

As noted in Section 3.2, out of 80 subjects who participated in the experiment, 24 were removed from the analysis: 11 subjects were removed due to non-monotone choices (assuming transitivity) in at least one pair of questions, 7 due to cyclic choices between $\{R_1, Same, Diff\}$, 4 for disagreeing with the symmetry over colors and urns expressed in (2.1), and two more subjects in the first session were caught cheating and their choices were excluded from the analysis. Seven more subjects were removed from the analysis of CCE.[39] The goal of this Appendix is to note that the tendencies and associations highlighted in the body of the paper persist even if we weaken the inclusion criteria used in Section 3.2.

In order to include as many subjects as possible, several relaxations of the inclusion criteria were employed. First, the answer to the non-incentivized question concerning symmetry was ignored. Second, in case of cyclic choices among $\{R_1, Same, Diff\}$, the direct comparison between $Same$ and $Diff$ was not taken to invalidate the choices made in One vs Two (which relies on comparing R_1 to $Same$ and R_1 to $Diff$). Third, if a subject had non-monotone choices in only one pair of questions (assuming transitivity), (s)he was not removed from the analysis. Fourth, choices in other questions, together with transitivity, were used in order to extend the preferences to the missing direct comparison.[40] With all these adjustments, we were able to retain 77 subjects.[41]

The tables below replicate Tables 3.1, 3.2 and 3.3 for the larger group of subjects.

[39] Due to multiple switching points, extreme or missing answers, or research assistant's error in copying the switching point from Question 9 to Question 10.

[40] This applies to 5 subjects. An extreme example is provided by Subject 315: (s)he did not agree with the suggested symmetry in colors and urns, and her/his choices in R_1 vs $Same$ were inconsistent with monotone preferences. However, the choices in R_1 vs $Diff$ and $Diff$ vs $Same$ were consistent with $R_1 \succ Diff$ and $Diff \succ Same$, so this subject was classified as exhibiting $One \succ Two$.

[41] We omitted only the two subjects who were caught cheating and another subject who made choices inconsistent with monotone preferences in both R_1 vs $Same$ and R_1 vs $Diff$ and who had multiple switching points in CCE.

Ellsbergian Ambiguity	**One vs Two**			**Total**
	One$\succ$Two	**One$\sim$Two**	**One$\not\succsim$Two**	
Averse	24	11	16	51
Neutral	4	11	2	17
Seeking	1	2	3	6
Total	29	24	21	74
Fisher exact test (excluding non-monotone in Ellsberg) p-value=0.011<0.05				

Table D.1: Ellsbergian ambguity and One vs Two

	CCE	**not CCE**	**Total**
# of subjects	36	31	67
% of subjects	53.7%	46.3%	100

Table D.2: Correlation Certainty Effect

Same vs Different	**Ellsberg**		**Total**
	PS	**notPS**	
Same $\sim$ Diff	14	17	31
Same $\nsim$ Diff	2	39	41
Total	16	56	72
p-value Fisher exact test = 0.00008			

Table D.3: Ellsbergian ambiguity and Same versus Different

The marginal distributions of ambiguity attitude expressed in the standard Ellsberg experiment[42] and in One vs Two are remarkably similar to those for the smaller sample. The ambiguity attitudes in both behaviors are significantly associated but distinct. Slightly more than half of the subjects exhibit CCE.[43] As noted for the smaller sample, almost 70% of the subjects who were not probabilistically sophisticated in the Ellsberg questions were not indifferent between *Same* and $Diff$.[44]

[42] 3 subjects made choices inconsistent with monotone preferences in the Ellsberg questions.

[43] 10 subjects were removed from the analysis of CCE: 5 had extreme switching points (1, 99 or 100); 3 did not answer question 10 in which they were asked to compare f_1 to g_1; 2 had the wrong $\bar{x}$ from question 9 inserted in question 10.

[44] Almost 77% of them strictly preferred *Same* to $Diff$. Two subjects made choices inconsistent with monotone preferences in *Same* vs $Diff$.

E Experimental Instructions

This section contains the text of the instructions.[45]

Each of the two jars (Jar #1 and Jar #2) contains 10 marbles. Each marble is either green or blue. The number of green (and blue) marbles in each jar is unknown – it could be anything between 0 and 10. The two jars may contain different numbers of green (and blue) marbles.
At the end of the experiment, one marble will be drawn from each jar.

Each of the 10 questions below offers you a choice between bets on the colors of the 2 marbles that will be drawn at the end of the experiment. One of the questions will be selected at random according to the protocol specified in the following paragraph, and your chosen bet in that question will determine your payment. For example, suppose that in the question that was selected for payment you choose the bet "$100 if the marble drawn from the Jar #1 is green, otherwise $0". If the marble drawn from Jar #1 is indeed green – you will win $100, and if it is blue – you will win nothing (both are in addition to the payment of $10 you received for arriving to the experiment on time).

To select the question that will determine your payment, participants will be divided into two groups. One participant from each group will be randomly selected and will roll 3 dice for each participant in the other group: a 10-sided die that produces a number between 1 and 10, and two 10-sided dice that produce a number between 1 and 100. They will write the two numbers on notes that will be folded and inserted into sealed envelopes distributed among participants in the experiment. The first number will be used to select the question that will determine your payment. In case question 9 (which includes many sub-questions) is selected by the first die, the second number will be used to select the sub-question that will determine your payment. Do not open the envelope you receive until you complete answering all the questions and you are told to open it. Remember that the question is chosen before you make any choices.

This protocol of determining payments suggests that you should choose in each question as if it is the only question that determines your payment.

Remember that the compositions of both jars are unknown, so it does not matter if a bet is placed on a green or a blue marble. Similarly, it does not

[45] The original instructions were formatted in MS-Word and are available upon request.

matter if a bet is placed on Jar #1 or #2. Below are some examples that demonstrate this principle:

- "$100 if the marble drawn from the Jar #1 is *green*" and "$100 if the marble drawn from the Jar #1 is *blue*" are equally good.
- "$100 if the marble drawn from the Jar *#1* is green" and "$100 if the marble drawn from Jar *#2* is green" are equally good.
- "$100 if both marbles drawn are *green*" and "$100 if both marbles drawn are *blue*" are equally good.
- "$100 if the marble drawn from the Jar #1 is green and the marble drawn from the Jar #2 is blue" and "$100 if the marble drawn from the Jar #1 is blue and the marble drawn from the Jar #2 is green" are equally good.

Do you agree that the two bets in each pair are equally good? YES NO (circle one)

Before choosing between bets please choose a fixed color (green or blue) and a jar (#1 or #2) for which you will be paid if you choose certain bets in the questions below. For example, in question 1 you can choose to be paid if the marble drawn from Jar #1/#2 is green/blue. Note that you must make the same choice for all the questions below.
Please circle and choose your set jar and color:
Your fixed jar: #1 / #2
Your fixed color: green / blue
The choice of jar and color will apply to bets 1, 3, 5, 7, 13 and 15 below.

Question 1 (circle 1 or 2)

1. \$100 if the marble drawn from the fixed jar is of the fixed color
2. \$101 if the two marbles drawn are of different colors (one green and one blue)

Question 2 (circle 3 or 4)

3. \$101 if the marble drawn from the fixed jar is of the fixed color
4. \$100 if the two marbles drawn are of different colors (one green and one blue)

Note: Bets 1 and 3 pay under the same conditions but Bet 3 offers more money if you win (\$101) than Bet 1 (only \$100). Therefore anyone who prefers to earn more money would view Bet 3 as better than Bet 1. Similarly, Bets 2 and 4 pay under the same conditions but Bet 2 pays more money if you win than Bet 4. Therefore anyone who prefers to earn more money would view Bet 2 as better than Bet 4. If in one of the questions you choose the bet that pays \$100, it makes sense that in the other question you choose the corresponding bet. This follows since the corresponding bet pays \$101 (instead of \$100), and the payment to the alternative bet decreases from \$101 to \$100. *Please review your choices in questions 1 and 2 in light of this logic.* Notice that identical logic applies to the other questions (3-4, 5-6, 7-8).

Question 3 (circle 5 or 6)

5. \$100 if the marble drawn from the fixed jar is of the fixed color
6. \$101 if the two marbles drawn are of the same color (two greens or two blues)

Question 4 (circle 7 or 8)

7. \$101 if the marble drawn from the fixed jar is of the fixed color
8. \$100 if the two marbles drawn are of the same color (two greens or two blues)

Question 5 (circle 9 or 10)

9. \$101 if the two marbles drawn are of the same color (two greens or two blues)

10. \$100 if the two marbles drawn are of different colors (one green and one blue)

Question 6 (circle 11 or 12)

11. \$100 if the two marbles drawn are of the same color (two greens or two blues)

12. \$101 if the two marbles drawn are of different colors (one green and one blue)

I will now fill an empty third jar (#3) with 5 green and 5 blue marbles. The following two questions ask you to choose between a bet on the color of a marble drawn from this jar and a bet on the set jar (#1 or #2) and set color.

Question 7 (circle 13 or 14)

13. \$100 if the marble drawn from the fixed jar is of the fixed color

14. \$101 if the marble drawn from Jar #3 (that is known to contain 5 green and 5 blue marbles) is green.

Question 8 (circle 15 or 16)

15. \$101 if a marble drawn from the fixed jar is of the fixed color

16. \$100 if a marble drawn from Jar #3 (that is known to contain 5 green and 5 blue marbles) is green.

Question 9

Bet A pays $100 if the marble drawn from Jar #1 is green/blue (circle one) <u>and</u> the marble drawn from Jar #2 is green/blue (circle the other color).
Bet B pays $x if the two marbles drawn are of different colors.

Before you choose between the two bets above, you must know the value of x. For example, if x=100, then you will probably choose Bet B. The rationale behind this is that if you win with Bet A, then you will also win with Bet B, but there are cases in which only Bet B wins. Similarly, if x=0, then you will probably choose Bet A since it alone provides some chance of winning money.

Below, you are asked to choose between Bet A and Bet B for each value of x indicated in the list below (note that the list is on two pages). Note that while Bet A does not change between the lines, the amount paid in Bet B increases as you move down the list. Therefore, if you choose B on some line, it makes sense to choose B in every subsequent line.

If this question is chosen to determine your payment and if the relevant line was chosen (according to dice rolled by the two participants in the beginning of the experiment), then your payment will depend on the bet you choose. Therefore, you should make the choice in every line as if this is the only choice that will determine your payment in the experiment.

Remember that Bet B pays the amount specified on the line (between $1 and $100) if the two marbles drawn are of different color. Therefore, you will be paid if the marbles are as you specified for Bet A, but also if the colors of the two marbles are reversed.[46]

Line	**Bet A**	**Bet B: the value of x**	**Chosen Bet (circle A or B)**	
1	$100	$1	A	B
2	$100	$2	A	B
3	$100	$3	A	B
⋮	⋮	⋮	⋮	⋮
98	$100	$98	A	B
99	$100	$99	A	B
100	$100	$100	A	B

[46] The table in the experiment had 100 lines. Question 10 was not available to the subjects when they answered Question 9.

Question 10 (circle 17 or 18)

17. Pays according to Bet A in Question 9 or ____$[47] if the two marbles drawn are of the same color (either both green or both blue).

18. Pays ____$ for sure.

Reminder:
Bet A in Question 9 pays $100 if the marble drawn from Jar #1 is green/blue and the marble drawn Jar #2 is green/blue
(see question 9 for your choice of colors).

References

[1] M. Abdellaoui, A. Baillon, L. Placido and P.P. Wakker, The rich domain of uncertainty: source functions and their experimental implementation, *Amer. Econ. Rev.* 101 (2011), 695-723.

[2] N. Al Najjar and L. De Castro, Parametric representation of preferences, 2010.

[3] M. Amarante, Y. Halevy and E. Ozdenoren, Two-stage probabilistic sophistication, 2013.

[4] G. Aryal and D. H. Kim, A point decision for partially identified auction models, *J. Bus. Econ. Statist.* 31 (2013), 384-397.

[5] A. Baillon, Y. Halevy and C. Li, Incentive compatability of the random incentive system in ambiguity elicitation, 2013.

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

[7] W. Brock and S. N. Durlauf, Growth empirics and reality, *The World Bank Review* 15 (2001), 229-272.

[8] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci and L. Montrucchio, Ambiguity and robust statistics, *J. Econ. Theory* 148 (2013), 974-1049.

[47] Research assistants filled in the highest line in Question 9 on which the participant chose Bet A.

[9] S. H. Chew and J. Sagi, Event exchangeability: probabilistic sophistication without continuity or monotonicity, *Econometrica* 74 (2006), 771-86.

[10] S. H. Chew and J. Sagi, Small worlds: modeling attitudes towards sources of uncertainty, *J. Econ. Theory* 139 (2008), 1-24.

[11] I. Couso, S. Moral and P. Walley, Examples of independence in imprecise probabilities, in: Proceedings of ISIPTA'99, Ghent, Belguim, 1999, 121-30.

[12] B. De Finetti, La prevision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincare* 7 (1937), 1-68. English translation in *Studies in Subjective Probability*, 2nd edition, H.E. Kyburg and H.E. Smokler eds., Krieger Publishing, Huntington NY, 1980, pp. 53-118.

[13] A. P. Dempster, Upper and lower probabilities induced by a multi-valued mapping, *Ann. Math. Statist.* 38 (1967), 325-339.

[14] A. P. Dempster, A generalization of Bayesian inference, *J. Royal Stat. Soc.* Series B 30 (1968), 205-232.

[15] J. Dow and S.R. Werlang, Uncertainty aversion, risk aversion, and the optimal choice of portfolio, *Econometrica* 60 (1992), 197-204.

[16] J. Eichberger, J. Oechssler, and W. Schnedler, How do subjects view multiple sources of ambiguity? U. Heidelberg, 2013.

[17] K. Eliaz and P. Ortoleva, Multidimensional Ellsberg, Cal Tech, 2012.

[18] L.G. Epstein and M. Schneider, Ambiguity and asset markets, *Annual Rev. Finan. Econ*. 2 (2010), 315-346.

[19] L. G. Epstein and K. Seo, Symmetry of evidence without evidence of symmetry, *Theor. Econ.* 5 (2010), 313-368.

[20] L. G. Epstein and K. Seo, Bayesian inference and non-Bayesian prediction and choice: foundations and an application to entry games with multiple equilibria, 2013.

[21] H. Ergin and F. Gul, A theory of subjective compound lotteries, *J. Econ. Theory* 144 (2009), 899-929.

[22] L. Garlappi, R. Uppal and T. Wang, Portfolio selection with parameter and model uncertainty: a multi-prior approach, *Rev. Finan. Stud.* 20 (2007), 41-81.

[23] I. Gilboa and D. Schmeidler, Maxmin expected utility with non-unique prior, *J. Math. Econ.* 18 (1989), 141-153.

[24] B. Greiner, An online recruitment system for economic experiments. In: K. Kremer, V. Macho eds. Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63, Gottingen : Ges. fr Wiss. Datenverarbeitung, 79-93, 2004.

[25] Y. Halevy, Ellsberg revisited: an experimental study, *Econometrica* 75 (2007), 503-536.

[26] P. A. Haile and E. Tamer, Inference with an incomplete model of English auctions, *J. Pol. Econ.* 111 (2003), 1-51.

[27] E. Karni and Z. Safra, 'Preference reversal' and the observability of preferences by experimental methods, *Econometrica* 55 (1987), 675-685.

[28] G. King, Proper nouns and methodological propriety: pooling dyads in international relations data, *Internat. Organization* 55 (2001), 497-507.

[29] P. Klibanoff, M. Marinacci and S. Mukerji, A smooth model of decision making under ambiguity, *Econometrica* 73 (2005), 1849-1892.

[30] P. Klibanoff, S. Mukerji and K. Seo, Relevance and symmetry, 2011.

[31] D. Kreps and E. Porteus, Temporal resolution of uncertainty and dynamic choice theory, *Econometrica* 46 (1978), 185-2000.

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

[33] M. Machina and D. Schmeidler, A more robust definition of subjective probability, *Econometrica* 60 (1992), 745-780.

[34] U. Segal, The Ellsberg paradox and risk aversion: an anticipated utility approach, *Int. Econ. Rev.* 28 (1987), 175-202.

[35] U. Segal, Two-stage lotteries without the reduction axiom, *Econometrica* 58 (1990), 349-377.

[36] K. Seo, Ambiguity and second-order belief, *Econometrica* 77 (2009), 1575-1605.

[37] E. Tamer, Partial identification in econometrics, *Annual Rev. Econ.* 2 (2010), 167–195.

[38] E. Tamer, Incomplete simultaneous discrete response model with multiple equilibria, *Rev. Econ. Stud.* 70 (2003), 147-165.

[39] A. Tversky and C. R. Fox, Weighing risk and uncertainty, *Psych. Rev.* 102 (1995), 269-283.

[40] A. Tversky and P. P. Wakker, Risk attitudes and decision weights, *Econometrica* 63 (1995), 1255-1280.

[41] P. Walley and T.L. Fine, Towards a frequentist theory of upper and lower probability, *Ann. Statist.* 10 (1982), 741-761.