COMPARING AMBIGUOUS URNS WITH DIFFERENT SIZES*

EMEL FILIZ-OZBAY†, HUSEYIN GULEN‡, YUSUFCAN MASATLIOGLU§,
AND ERKUT Y. OZBAY*

Abstract. We use two-color Ellsberg urns to compare different ambiguous processes. By providing symmetric information on urns with different numbers of beads and keeping the most optimistic and pessimistic possibilities the same, we elicit subjects' preferences for the size of an ambiguous urn. Subjects prefer the bets from the ambiguous urns with more beads. Ambiguity averse subjects mainly drive this effect. We study the restrictions that our findings impose on the existing ambiguity models. Finally, we verify our experimental result with a mutual fund data.

Keywords: Ambiguity, Risk, Ratio Bias, Ellsberg's Experiment.

1. Introduction

In a two-color urn thought experiment of Ellsberg (1961), a decision maker (DM) prefers betting on an urn with 5 Black and 5 White beads (the risky urn) rather than an urn with a total of 10 Black and White beads with an unknown composition (the ambiguous urn). Following Ellsberg's thought experiment, normatively or prescriptively appealing theories of ambiguity emerged. The well-known ambiguity models include (i) multi-prior approach by Gilboa and Schmeidler (1989), (ii) Choquet expected utility by Schmeidler (1989), and (iii) models with second-order beliefs such as Ergin and Gul (2009), Klibanoff et al. (2005), Neilson (2010), and Seo (2009) (see also Machina and Siniscalchi (2014) for a survey of ambiguity models). The common prediction of these ambiguity aversion models is that a DM prefers betting on a less ambiguous urn to betting on a more ambiguous

*We are thankful to the Michigan Institute for Teaching and Research in Economics (MITRE) for the generous funding used for subject payments. The research project was initiated and experiments were conducted while Emel Filiz-Ozbay and Erkut Y. Ozbay were visiting University of Michigan, Department of Economics. The project was approved by University of Maryland and University of Michigan IRBs. We thank Yan Chen for allowing us to utilize the experimental laboratory in the School of Information at the University of Michigan. We also benefitted from fruitful discussions with Evan Calford, Yoram Halevy, Neslihan Uler, Michael Woeppel, and Frank Yates as well as the seminar participants at NYU-Economics, GSU-J. Mack Robinson College of Business, ESA-2016 Meetings, and SEA-2016 Meetings.

† University of Maryland, Department of Economics. E-mail: filizozbay@econ.umd.edu.
‡ Purdue University, Krannert School of Management. E-mail: hgulen@purdue.edu.
§ University of Maryland, Department of Economics. E-mail: yusufcan@umd.edu.
* University of Maryland, Department of Economics. E-mail: ozbay@umd.edu.

urn. Ellsberg's thought experiment has been widely confirmed in numerous laboratory experiments (see e.g., Camerer and Weber (1992) and Machina and Siniscalchi (2014) for detailed surveys).

Now, consider a DM who decides to place a bet between two ambiguous Ellsberg's urns, one with 2 beads and another with 1000 beads, for which the composition of Black and White beads in either urn is unknown. Other than the total number of beads in each urn, no information about the urns is provided to the DM. Does the DM prefer the urn with 2 beads, 1000 beads, or is she indifferent between them? Since the well-known models are not crafted to make any prediction regarding the preference for the urn size, any preference is compatible with the theory: if the DM perceives no difference in the amount of ambiguity between these urns, then she will be indifferent between the two, but if she perceives one bet as more ambiguous than the other, then she will prefer betting on the less ambiguous bet. For example, as in the example of Einhorn and Hogarth (1985), say the colors of the beads are determined by throwing them into two adjacent black and white paints from a distance of 20 feet. Given this process, the composition with all black beads or all white beads may be more "probable" in a 2-bead urn than in a 1000-bead urn, i.e. a DM may think that there must be at least some beads with the color she had bet on among those 1000 beads. Hence, an ambiguity averse DM may exhibit a preference for a larger urn. Alternatively, the amount of ambiguity may be an increasing function of the number of possibilities in an uncertain situation (see Einhorn and Hogarth (1986)). While in a risky urn, there is only one possibility, 1 black and 1 white bead, in an ambiguous urn with 2 beads, there are 3 possibilities, and in an ambiguous 1000-bead urn, there are 1001 possibilities. It may be harder to contemplate 1001 possibilities than 3 possibilities, so a larger urn may be perceived as more ambiguous. Hence, an ambiguity averse DM may exhibit a preference for a smaller urn. In this paper, we repeat Ellsberg's two-urn experiment by varying the number of beads in an urn. Since the urn size is observable, if it is playing a role in the decision of a DM, the size component, as a measure of ambiguous state space or complexity of the source of ambiguity, can be incorporated into the ambiguity models.

In many situations, individuals decide among different ambiguous situations where the size may have an effect in the amount of ambiguity. For example, imagine a DM selecting a day laborer outside a home improvement retailer, where workers congregate, for a job that does not require many qualifications (see Valenzuela Jr (2003) for the U.S. day laborer market). Suppose that each worker is either good or bad. The DM does not know the distribution of workers' types, and she will pick the first worker in the line at the retail location. If there are two retail locations such that more workers gather at one than the other, would the size of the crowd (as the size of an Ellsberg urn) matter for the DM even if the size of the crowd did not indicate anything about the chance of getting better service? Alternatively, consider a company that offers a prize to lucky winners who are determined by chance. Say, you are in a situation similar to Charlie in "Willy Wonka & the Chocolate Factory." You are planning to buy a Wonka Bar some of which include a Golden Ticket that gives to the finders a full tour of the factory as well as a lifetime supply of chocolate. There are two stores: a small store that carries few Wonka Bars and a supermarket that carries a lot of Wonka Bars. Do you prefer to purchase a Wonka Bar from the small store or from the supermarket?

Our experiments explicitly address this type of question in a context-free environment, focusing on the size controlling all other possible effects. In our experiments, subjects compare two ambiguous urns containing a different number of beads (2, 10, or 1000). We also re-conduct Ellsberg's two-urn experiment where a risky and an ambiguous urn are compared by varying the total number of beads in each urn. Our results indicate that there is a preference for larger size when comparing ambiguous urns; and the preference for the larger urn is mainly driven by ambiguity averse subjects.

Based on our experimental findings, we revisit the existing models and discuss what our results impose on these models. For example, in the Maxmin Expected Utility model of Gilboa and Schmeidler (1989), the DM evaluates the alternatives based on multiple priors and chooses with respect to the worst-case scenario. We introduce the concept of "plausibility" such that the set of priors when evaluating a smaller urn should include more pessimistic priors than the set of priors when evaluating a larger urn so that the ambiguity averse DM will prefer a larger urn over a smaller one. Another conceivable way of modeling the decision problem could be as follows: the DM views an ambiguous urn as a two-stage lottery in a compound lottery sense. In the first stage, the composition of beads (i.e. number of Black and White beads) in an n-bead urn is determined. In the second stage, a bead is drawn from the urn, which has an objective distribution of colors in this stage. In this view, the first stage involves ambiguity, but the second stage does not.[1] We provide a sufficient condition for the first stage belief formation so that the smooth ambiguity model (Klibanoff et al. (2005)) defined on two-stage lotteries can accommodate our data.

Finally, for the external validity of our results, we conduct an empirical investigation to see whether investors take the number of assets held by a mutual fund (a measure of size) into account when deciding to allocate capital across different funds. Mutual funds have

[1]This framework is similar to Segal (1987) where ambiguous prospects are analyzed as lotteries with two stages. He relaxes the Reduction of Compound Lotteries axiom.

been a preferred choice of researchers when testing the predictions of ambiguity aversion in a non-experimental setting since the documented lack of predictability in mutual fund returns may make the investor consider each fund as an ambiguous process.[2] If the commonly used determinants of fund performance (such as fund flows, return volatility, fund fees, fund age, total net assets of the fund) of two funds are the same, does the investor have a preference for the fund with more asset holdings? This choice environment shares significant similarities with our two-urn experiment. In this non-experimental setting, we provide evidence consistent with the findings of our experiment. Controlling for commonly used determinants of fund flows, which include lagged raw and benchmark-adjusted returns, lagged fund flows, and total net assets as well as controls such as fund return volatility, fund fees, and fund age, we find that fund flows are positively and significantly related to the number of assets held by a fund. Moreover, when we repeat the above tests to predict future fund returns, we do not find a significant relation between the number of a fund's assets and future fund returns. These findings verify that in such an ambiguous choice environment, investors prefer mutual funds with more holdings even though this does not result in better investment performance.

Closely Related Experimental Literature: There are other ambiguity experiments that require subjects to compare different ambiguous processes. Halevy and Feltkamp (2005) and Epstein and Halevy (2018) detect a lack of confidence in (or uncertainty about) the joint distribution of two ambiguous Ellsberg urns. Abdellaoui et al. (2011) highlight the fact that probabilistic sophistication is usually violated when different sources of ambiguity are compared, even though it is satisfied within a single source. Our experiment complements this literature since in the comparison of two ambiguous urns with different sizes, there is no ambiguous correlation between the urns or asymmetric familiarity towards one urn. Additionally, Chew et al. (2017) elicit the subjects' preferences for partial ambiguity. Chew et al. (2017) observe aversion to an increased number of possible compositions for their "interval" and "disjoint" ambiguity processes. However, unlike our setup, the information given on two ambiguous processes is not symmetric in Chew et al.'s (2017) setup. Particularly, the most optimistic and pessimistic states of the world are different when the partial ambiguity level changes in the Chew et al. (2017) experiment, and that is not the case in our ambiguity comparisons. Therefore, their findings are not directly applicable to the decision problems we pose.

In a typical ambiguity experiment, the urn size is fixed and subjects are asked to compare risky and ambiguous urns. Pulford and Colman (2008) repeated the standard Ellsberg experiment varying the urn sizes. They find that the percentages of subjects choosing

[2] See, for example, Antoniou et al. (2015), Li et al. (2016).

the risky urn are not significantly different for the urns containing 2, 10 or 100 balls and conclude that the ambiguity attitude is not affected by the urn size.[3] Although our focus is to understand the preference between two ambiguous urns with different sizes, we also ask the standard Ellsberg questions when varying the size. We confirm their findings for 2- and 10-bead urns. However, when the size of the urn is very large, particularly 1000 - a size they do not investigate-, we find that the percentage of subjects choosing the risky urn is significantly smaller. We argue that due to our main findings of "preference for larger urn under ambiguity," it may be misleading to make inferences about the robustness of ambiguity attitudes only looking at the robustness of preferences between risky and ambiguous urns with varying sizes. Particularly, the larger urns may be perceived to be less ambiguous.

The paper is organized as follows. Section 2 describes our experimental design, and Section 3 presents the results. Section 4 discusses some existing models in the context of our decision problems. Section 5 includes an empirical analysis that checks the findings of our experimental findings in a non-experimental setting: mutual fund investment problem. Section 6 concludes. The instructions for the experiment are presented in the Appendix.

2. Experimental Design and Procedures

The experiments were conducted at the Experimental Laboratory of the School of Information, University of Michigan where 120 University of Michigan students participated.[4] The sessions lasted approximately 40 minutes. Subjects were paid in cash at the conclusion of the experiment and average earnings were approximately \$24 (including a \$7 participation fee). This experiment was conducted with pencil and paper.

Before the experiments started, we prepared six urns filled with Black and White beads. Table 1 summarizes the contents of each urn. The urns R2, R10, and R1000 involved only risk where half of the beads in these urns were Black and half of the beads were White. These urns contained 2, 10, and 1000 beads in total, respectively. The subjects were informed about the exact content of these urns and had the chance to check the urns

[3] One should be careful interpreting the evidence in Pulford and Colman (2008). We argue that there are two major flaws in their design. First, there is no ambiguity. Instead, they elicit the preferences between a compound lottery (which they interpret as an ambiguous urn) and its reduced form version (which they interpret as a risky urn). While Gillen et al. (2019) find high correlation between ambiguity neutrality and reduction of compound lotteries, Abdellaoui et al. (2015) illustrate significant differences between attitudes toward ambiguity and compounding. Second, choosing one alternative is interpreted as if it is a strict preference.

[4] Note that our subject size is comparable to sample size of a single treatment of other ambiguity experiments in the literature. For example, Crockett et al. (2019) have 122 subjects in their largest treatment, Halevy (2007) has 104 subjects in the main treatment, Epstein and Halevy (2018) have 153 subjects in the largest treatment, Abdellaoui et al. (2015) have 115 subjects in their largest treatment.

to make sure that they understood the objective probability of drawing a black or white bead. The urns A2, A10, and A1000 were the ambiguous ones. These urns also contained 2, 10, and 1000 beads in total, respectively. The subjects were told the total number of Black and White beads in each of these ambiguous urns, but not the exact number of Black or White beads in any of them. We also did not tell them the procedure we used to fill these urns until the end of the experiment. Moreover, the subjects were not allowed to check the content of these urns. All urns were placed on a tall desk, which could be seen by each subject clearly, and once the experiment started no one (even the experimenter) could touch the urns.

Table 1. Content of the Urns Used in the Experiment

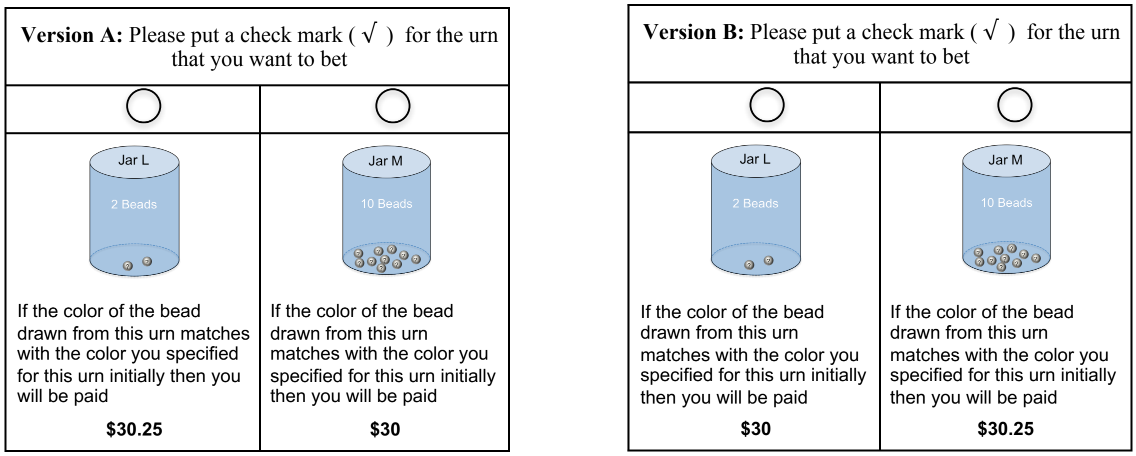
Urn	# of Black Beads	# of Total Beads	Distribution Information Given to Subjects
R2	1	2	Yes
R10	5	10	Yes
R1000	500	1000	Yes
A2	Unknown to the subjects	2	No
A10	Unknown to the subjects	10	No
A1000	Unknown to the subjects	1000	No

At the beginning of a session, each subject signed an informed consent form and received the written instructions provided in the Appendix. Subjects were given time to read the instructions and then an experimenter read the instructions aloud as well.

We asked subjects to make binary comparisons between bets on these urns. We chose this well accepted methodology to avoid complications introduced by other mechanisms such as eliciting cardinal preferences.[5] They were asked seven binary decision problems, and there were two versions of each problem (Versions A and B), as explained below. Hence, the subjects made 14 decisions in total. We paid the subjects only for one decision they made, and the paying decision was determined before the subjects made the decisions.[6] In order to determine the paying problems, the experimenter rolled a die, noted the outcome on a piece of paper, and put it in a sealed envelope. Then the envelopes were distributed to

[5]We elicited ordinal preferences rather than cardinal valuations of the bets by using methods such as Becker et al. (1964), which is used in Halevy (2007), or a choice list method used in Abdellaoui et al. (2011) and Abdellaoui et al. (2015). These methods introduce other complexities that we wanted to avoid.

[6]Random incentive system is the commonly used mechanism to prevent subjects from hedging over the randomization between problems. The incentive compatibility of this mechanism has been recently challenged (see e.g. Bade (2015)). Baillon et al. (2015) provide the sufficient conditions for this to be incentive compatible and it is used in Epstein and Halevy (2018), and Loomes et al. (1989). The order of the resolution in our design was made in advance to satisfy the sufficient conditions and it is identical to Loomes et al. (1989) and Epstein and Halevy (2018).



(A) Version A (B) Version B

Figure 1. Versions of a Sample Decision Problem

the subjects. The subjects knew that the paying decision problem had been determined before they made decisions, and they knew that they would learn the paying decision problem after the experiment was finalized. This prevented subjects from hedging over the randomization between problems.

After we introduced the six urns summarized in Table 1, we asked subjects to pick a color to bet on for each urn. The purpose of the selection was twofold: (i) we wanted to convince the subjects that the experimenters did not have any bias toward a particular color, and (ii) we wanted them to have the same bet for an urn when that urn was presented in different decision problems to avoid hedging (see Epstein and Halevy (2018) for further discussion). The selection of colors was entered by the subjects.

A typical alternative in a binary decision is a bet on an urn that pays a positive prize if the initially selected color of the subject for this urn matches the color of the randomly drawn bead from the urn at the end of the experiment. If the subject's selection of color and the experimenter's draw for that urn do not match, then the subject receives zero from this bet. We used $30 or $30.25 as the prize for a bet. Figure 1 presents an example of the two versions of a decision problem.

In this sample decision problem, the decision maker chooses between an ambiguous urn with two beads (A2) and an ambiguous urn with ten beads (A10). We elicit the preference of the decision maker between A2 and A10.[7] The only difference between the two versions is that urn A2 pays $30 in one version and $30.25 in the other. Similarly, the prizes for urn

[7]We labeled the urns with letters L, M, N, P, R, and S rather than A2, R2, etc. in the experiment. Moreover, we called them jars rather than urns.

A10 changes in two versions. If a subject chooses A2 over A10 in both versions, given that subjects prefer more money, it is reasonable to interpret this behavior as A2 being strictly preferred to A10. On the other hand, if she prefers A2 in version A and A10 in version B, then she chooses the urn with \$30.25 all the time. This might be because she thinks that her chances of winning are the same on these two urns; hence the prize determines her choice. Such a subject is always expected to pick the higher prize urn even if the prize difference was smaller than \$0.25. It might also be the case that she actually thinks her chance of winning is higher on one urn, say A2, but \$0.25 additional prize cancels this out, and she chooses A10 in version B. Such a subject would pick A2 if the prize difference was small enough not to cancel out the likely effect. We believe that \$0.25 was a small enough prize difference to elicit strict preferences of most subjects between urns with different sizes. Therefore, observing a subject choosing the urn with a higher prize in both versions does not reveal anything about the preferences between urns. Finally, if a subject picks the urn that pays \$30 in both versions, she leaves money on the table.

Table 2 reports the seven binary comparisons presented to the subjects. Remember that there were two versions of each comparison. The first two rows of Table 2 test whether there is any preference for the size of ambiguous urns. Rows three and four test ratio bias when pure risk is involved. Finally, the last three rows test whether the preference for a risky urn when compared with an ambiguous one is robust to the size of the urn. We presented the decision problems in different orders to the subjects to control for the potential order effect. We had four different orders of the decision problems with a balanced number of subjects in each. In each ordering, versions A and B of the same decision problem were presented on the same page; three ambiguity attitude problems were always asked the last (as they are not our main decision problems). In two orderings we first elicited preferences for the size of ambiguous urns, then elicited preferences for the size of risky urns. In the other two orderings we switched this order. We randomized which urn was presented on the left in a decision problem; hence, it was not the case that always a larger urn or always a smaller urn was presented on the left hand side. Note that the subjects were allowed to choose whichever order they wanted to answer questions as all the decision problems were handed in the same package.

We included the urn with ten beads because that was the size of urns typically used in ambiguity experiments (see for example, Halevy (2007) and Epstein and Halevy (2018)). We chose size two as well because it is also used in the literature (Abdellaoui et al. (2015) and Epstein and Halevy (2018)) and it is the smallest even number where we can generate a 50% chance of winning for each color for the urns with pure risk. We used size 1,000 because in most of the psychology experiments on ratio bias phenomena, they reported

Table 2. Decision Problems Used in the Experiment

The problems measure			
Preferences for Size under Ambiguity	A2	vs.	A10
	A10	vs.	A1000
Preferences for Size under Pure Risk	R2	vs.	R10
	R10	vs.	R1000
Ambiguity Attitudes	A2	vs.	R2
	A10	vs.	R10
	A1000	vs.	R1000

frequencies to subjects in a sample of 1,000 (Barratt et al. (2005); Pinto-Prades et al. (2006)).

After the subjects made their choices on the binary decision problems, the experimenters collected their choice sheets and then drew the beads from the urns while the subjects observed. After that, the subjects were allowed to open the sealed envelopes to see the decision problem for which they would be paid. Each subject met the experimenter individually to find out what she had chosen for her paying decision problem and whether the color she initially had bet on for her chosen urn in that problem matched the color of the bead drawn from that urn. If they matched, she received the specified prize in addition to her participation fee of \$7. Otherwise, she received only the participation fee.

3. Experimental Results

Out of 120 subjects, three subjects were excluded from the following analysis because they chose the smaller prize in both versions of the decision problem, and one subject was excluded for not answering all of the questions.[8] Hence, the analyses are based on 116 subjects.[9]

We start with reporting the aggregate data on Table 3. Each column represents a set of binary questions. For example, the first column reports subjects' preferences when they compared urn A2 and urn A10. The cells that contain the majority of the subjects for a given decision problem are highlighted. The first two rows report the percentages of subjects who preferred larger and smaller urns, respectively, in the corresponding decision problem. Recall that if a subject chose to bet on the same urn in two versions of a

[8]Choosing the smaller prize in both versions of the decision problems implies that subject violates either monotonicity or transitivity.

[9]Our results do not change if we include them.

decision problem, we concluded that she has a strict preference for that urn. The third row presents the percentage of subjects who chose the urn with the higher prize ($30.25) in both versions of the corresponding decision problem. Even though those subjects who are indifferent between the two urns are reported in that row, being in that group does not imply indifference. Hence, our identification of strict preference is conservative.

Note that when ambiguous urns are compared (in the first two columns of Table 3), the majority of subjects strictly prefer the larger urn to the smaller urn (62.93% in A2 vs. A10 ($z = 74.136, p < 0.0001$) and 59.48% in A10 vs. A1000 ($z = 61.370, p < 0.0001$)). This evidence indicates that the underlying process generating ambiguous events matters, and the preference is toward larger urns even though we have utilized a conservative measure in this analysis. A few subjects preferred the smaller urn to the larger one (8.62% in A2 vs. A10 and 10.34% in A10 vs. A1000). Approximately one third of the subjects chose the higher prized urn under ambiguity.

Even though some subjects showed strict preferences for the urn size in pure risk questions, the majority of the subjects chose the urn with the higher prize in both versions of these decision problems, i.e. in the pure risk questions (R2 vs. R10 and R10 vs. R1000), the percentage of subjects choosing the higher prize is greater than those choosing larger urn and smaller urn (for all such comparisons $p < 0.0001$).

However, note that among the subjects in the first two rows, the majority picked the larger urn. Even though it is not reported in the table, recall that almost negligible percentages (3/119) of subjects chose the urns with the lower prize in both versions of a decision problem.

Table 3. Preferences for the Urn Size under Ambiguity and Risk, $N = 116$

Preferences for	**A2 vs. A10**	**A10 vs. A1000**	**R2 vs. R10**	**R10 vs. R1000**
Larger Urn	62.93%	59.48%	24.14%	31.03%
Smaller Urn	8.62%	10.34%	16.38%	12.93%
Higher Prize	28.45%	30.17%	59.48%	56.03%

Any decision making under risk procedure that is based on probabilities predicts indifference between urns R2, R10 and R1000. A violation of this prediction is identified as ratio bias in the literature (Kirkpatrick and Epstein (1992)). Psychologists define ratio bias as the tendency for people to judge a low probability event as more likely when presented as a large-numbered ratio, such as 20/100, than as a smaller-numbered but equivalent ratio, such as 2/10.[10] According to our elicitation method, those subjects who do not

[10]While controlling for ratio bias, our design approach differs from the psychology experiments in three ways. First, we incentivize the subjects for revealing their true preferences with methods used more

have ratio bias need to choose the higher prize in both versions of the decision problems that involve only risk. This is indeed the case for the majority of our subjects (59.48% on R2 vs. R10 and 56.03% on R10 vs. R1000 in Table 3). Hence, we observe less ratio bias than documented in the psychology literature. This might be because we use different methodology and incentivized the decisions rather than providing fixed rewards, which differs from the common practice in psychology experiments. It may also be due to the fact that the chance of winning is not small in our risky urns (the probability of winning is 0.5) but that probabilities are typically less than 0.2 in the literature (e.g. Denes-Raj and Epstein (1994)). The source of the difference is not within the scope of this project.

Next we restrict analysis to those subjects whose decisions are consistent with probability theory on our decision problems with pure risk. This exercise helps us see whether the preferences for the urn size under ambiguity come from ratio bias. Table 4 reports the preferences for the urn size under ambiguity for only those subjects who chose the urns with a higher prize on pure risk questions (there are 56 subjects presenting such behavior). If we interpret those subjects' behavior as indifference on the urn size when the probability of winning is 0.5, then these subjects are the ones who have a better understanding of the probability. Again a significant number of subjects (see Table 4) strictly preferred betting on the larger urns to betting on the smaller urns under ambiguity, i.e. when they compared both A2 vs. A10 and A10 vs. A1000 ($z = 31.033, p < 0.0001$; $z = 27.187, p < 0.0001$, respectively). This indicates that the result for preference for the larger urn under ambiguity is not driven by the subjects who fall into the ratio-bias.

Additionally, we check whether the preference for size under ambiguity is an amplification of the preference for size under risk. According to this amplification hypothesis, the biases are amplified as the environment becomes more uncertain.[11] In our setup, this hypothesis implies that individuals who prefer the larger urn under risk should prefer the larger urn under ambiguity; similarly, individuals who prefer the smaller urn under risk should prefer the smaller urn under ambiguity. However, the individuals who are indifferent in terms

---

frequently by economists than psychologists. Second, our subjects can reveal their indifference between alternatives to a certain degree. Third, our risky urns are filled with 50% Black and 50% White beads; hence, we do not have low probability events. To the best of our knowledge, the only paper on ratio bias in the economics literature is Lefebvre et al. (2011). They argue against ratio bias and conclude that its relevance for economic applications is questionable. They use a completely different methodology where subjects made sequential binary comparisons between urns. In their experiment, the small urn always offers 1−in−10 chance of winning the prize and the large urn starts with 10−in−100 chance of winning and gradually decreases to 9−in−100, 8−in−100, . . ., 3−in−100. They find that the percentage of subjects who prefer the larger urn when the probabilities are equal depends on which order this choice problem is presented. Their subjects have to express a strict preference between the two urns, which might be an issue, as we discuss in our design.

[11]For example, Maafi (2011) experimentally shows that the preference reversals under ambiguity are stronger than those under risk.

of urn size under risk may have a preference for urn size under ambiguity; hence, the preference of smaller and larger urns should increase under ambiguity rather than under risk. Unlike the prediction of the amplification hypothesis, independent of the preference for size under risk, there is a preference for a larger urn under ambiguity. For example, between the urn sizes of 2 and 10 beads, among the 19 subjects who chose the smaller urn under risk, only 4 of them preferred the smaller urn under ambiguity.

We also check whether subjects exhibit monotonic preference with respect to urn size under ambiguity. We find only 3 subjects violating such monotonicity. There is no subject violating transitivity in this group. In the overall subject pool, we observe a single transitivity violation.

Table 4. Preferences for the Urn Size: without Ratio Bias, $N = 56$

Preference for	**A2 vs. A10**	**A10 vs. A1000**	**R2 vs. R10**	**R10 vs. R1000**
Larger Urn	53.57%	46.43%	0%	0%
Smaller Urn	5.36%	3.57%	0%	0%
Higher Prize	41.07%	50.00%	100%	100.0%

Table 5 reports the percentages of the preferences in three sets of decision problems, R2 vs. A2, R10 vs. A10, and R1000 vs. A1000. The percentages of subjects choosing the risky urn are 74.14%, 73.28% and 63.79% in 2-, 10-, 1000-bead urns, respectively. The percentages of subjects preferring the risky urn on 2-bead urns and 10-bead urns are not significantly different ($p = 0.8820$). However, the percentage of subject preferring the risky urn on 1000 bead urns is significantly lower ($p < 0.05$).

In a typical experiment, this table is interpreted as the measure of ambiguity attitudes. The first row reports the percentage of subjects who preferred the urn with pure risk in both versions of the corresponding decision problem. Those are identified as ambiguity averse. The second row reports the percentage of subjects who preferred the ambiguous urn in both versions of the corresponding decision problem (i.e., lack of ambiguity aversion). The subjects in the third row are the ones who chose the urn with higher prize in both versions of the ambiguity attitude questions. Such a subject might have been indifferent between the two urns or the \$0.25 prize difference was not high enough to identify her strict preferences between the urns.[12] Importantly, existing experiments typically fix

[12]Nevertheless, we can say that even if she strictly prefers one to another, the intensity of her preference is minor and such a preference is economically irrelevant. For example, say a subject strictly prefers the risky urn in R2 vs. A2 but the higher prize in R1000 vs. A1000. This means that in 2-bead urns, this subject prefers the risky urn more than an addition of \$0.25, but in 1000-bead urn, even if she prefers one over the other, \$0.25 is valued more than expressing her preference. Therefore, the subjects who prefer the higher prize can be defined as "almost ambiguity neutral." Note also that an ambiguity averse subject with asymmetric beliefs on colors might be indifferent between ambiguous and risky urns. While this is

the number of beads, and the identification or the estimation is made without varying the number of beads. If we adopt the same identification strategy, in line with the existing ambiguity experiments, in our experiment a majority of the subjects are identified as ambiguity averse. Although the percentages of the ambiguity averse subjects are not different in 2-bead and 10-bead urns, the percentage decreases in a 1000-bead urn.[13] However, such a conclusion on changing ambiguity attitude based on urn size may be problematic. In order to reach a conclusion on ambiguity attitudes, the perceived ambiguity of an individual should stay the same across different sized urns (see e.g. Klibanoff et al. (2014) on ambiguity attitude and perceived ambiguity distinction). The results about ambiguity attitude measured by different sized Ellsberg urns should be viewed in light of our result of preference for a larger urn. We have already shown in our main result that there is a "preference for a larger urn under ambiguity," implying that the amount of ambiguity decreases in a larger urn. Perhaps the ambiguity attitudes of the subjects stay the same across different urns, but larger urns are perceived to be less ambiguous.

Table 5. Ambiguity attitudes, $N = 116$

Preference for	**R2 vs. A2**	**R10 vs. A10**	**R1000 vs. A1000**
Risky Urn	74.14%	73.28%	63.79%
Ambiguous Urn	10.34%	10.34%	12.07%
Higher Prize	15.52%	16.38%	24.14%

Finally, the majority of subjects (63 subjects) had a strict preference for the risky urn over the ambiguous one in all three versions of the ambiguity attitude problems. Among these subjects, 68.25% preferred the larger urn when they compared the two ambiguous urns with different sizes. Hence, when we identify such a consistent preference for a risky urn as ambiguity aversion, then the preference for larger urn under ambiguity that is reported in Table 3 is mainly driven by ambiguity averse subjects.

4. Incorporating the Size of the Urn into Ambiguity Models

It has been well-documented that the source generating ambiguity affects the decisions (see e.g. Tversky and Fox (1995)). However, as argued in Klibanoff et al. (2012), if this source is unknown, as in a typical Ellsberg experiment, any behavior of the DM is rationalizable (see also Epstein (2010)). In our experiments, a subject is not informed about how the

theoretically possible, we believe that subjects viewed the two colors symmetrically in our experiment. Whether subjects have asymmetric beliefs on colors is beyond the scope of this paper.

[13] In line with our result in 2- and 10-bead urns, Pulford and Colman (2008) find that the percentages of subjects choosing the risky urn are not different but they do not study when the size of the urn is 1000.

experimenter filled the ambiguous urns and whether there is any relation between the ways two ambiguous urns are filled. Therefore, the DM is equally informed about the source generating uncertainty for two different ambiguous urns in our experiment. One may argue that a subject is free to think about those ambiguous urns anyway she wants and may evaluate each urn independently. However, this line of thinking can lead to any type of preference between two ambiguous urns with different sizes without any prediction power. In contrast, our subjects presented a systematic preference for a larger ambiguous urn over a smaller one. Next, we will revisit some well known ambiguity models in light of our experimental findings and discuss the restrictions the observed behavior imposes on those theories.

4.1. **Two-Stage Models.** Two-stage models describe decision problems as two-stage lotteries where the uncertainty in the first stage is vaguely described, and the one in the second stage is objectively specified. Motivated by theories that incorporate the failure of compound risk reduction idea (see Becker and Brownson (1964), Segal (1987)), two stage models (such as Klibanoff et al. (2005) and Ergin and Gul (2009)) developed ambiguity theories where the decision maker has distinct preferences across the two stages of resolution of uncertainty.

Assume that a subject in our experiments views an ambiguous urn of size n as a two-stage procedure. In the first stage, the experimenter generates the urn, and in the second stage a bead is randomly drawn from the urn. The state space of the first stage has $n+1$ events, one for each combination of black and white beads. Let $\{0B/nW, 1B/(n-1)W, \ldots, nB/0W\}$ be the state space of the first stage of the procedure that carries ambiguity. Figure 2 represents the two-stage procedure for a bet on a Black bead from an urn of size n. Note that the lotteries on the second stage have objective probabilities for the possible outcomes of zero and \$30. The question is what the decision maker thinks about the first stage of the procedure, i.e. how the experimenter created the urns. Let the decision maker face an urn with n beads and believe that the probability of having i Black beads in the urn is p_i^n. We use the superscript n because it is conceivable that this probability depends on the urn size the subject faces: Different urns might induce different beliefs. In our setup, the urns differ only in terms of their size.

Any experiment in which the same urn is used across decision problems (including Ellsberg's original experiment) is not capable of illustrating the effect of p_i^n. However, when different ambiguous urns are compared as in our experiment, the formation of p_i^n is the heart of the question. Being silent on how p_i^n is formed, the predictive power of the existing two-stage models diminishes. To illustrate this point, we focus on the smooth ambiguity model of Klibanoff et al. (2005).

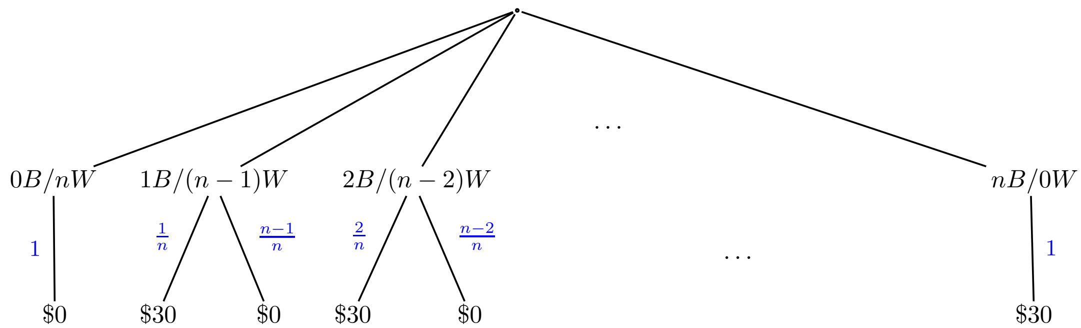


Figure 2. A Two-stage Compound Lottery

The smooth ambiguity model of Klibanoff et al. (2005) in our setting evaluates the uncertainty described in Figure 2 as

$$\sum_{i=0}^{n} \phi[EU(iB/(n-i)W)]p_i^n$$

where ϕ determines the ambiguity attitude and

$$EU(iB/(n-i)W) = \frac{i}{n}u(30) + \frac{n-i}{n}u(0)$$

represents the expected utility assigned to each node in stage 2. We can simplify the two stage lottery of Figure 2 to a single stage one as in Figure 3.

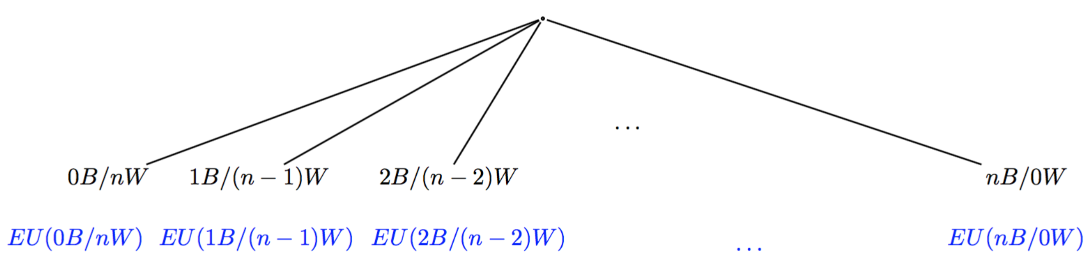


Figure 3. Reduced lottery: Expected utilities for the second stage lotteries are plugged.

First, we note that without any restriction on how p_i^n changes with n, any behavior can be rationalized by this model. For example, consider two urns with 2 and 10 beads. Let q^n be the uniform distribution ($q_i^n = \frac{1}{n+1}$ for all i) and r^n be the extreme distribution on two extreme points ($r_0^n = r_n^n = \frac{1}{2}$). For any non-linear ϕ, the preferences of a decision maker holding q^2 and q^{10} are completely opposite the preferences of a decision maker holding q^2 and r^{10}. Hence, any preference for the urn size can be accommodated by this model.

We next describe reasonable examples of beliefs depending on the urn size that can be consistent with our experimental results. Note that ambiguity aversion of a DM is characterized by concavity of ϕ of the smooth ambiguity model. It is a well known prediction that when the DM compares two reduced lotteries as in Figure 3, ϕ is concave if and only if the DM prefers the second order stochastically dominating lottery.

We now describe some beliefs that make the lottery corresponding to the larger urn second order stochastically dominate the lottery corresponding to the smaller urn. Let us consider two distributions which assume symmetry of beliefs between the two colors. The first is the uniform distribution, i.e. $p_i^n = 1/(n+1)$. The second has a unique underlying distribution of colors to be used while generating the urns, i.e. $p_i^n = \binom{n}{i}\lambda^i(1-\lambda)^{n-i}$, where λ is the probability of a Black bead in the underlying distribution of colors. The first scenario is as if the experimenter picked a number from $\{0, \ldots, n\}$ by using the uniform distribution while deciding how many Black beads to put in the urn of size n. The second scenario is as if the experimenter used a binomial distribution with underlying probability λ and drew n beads using such a binomial distribution to generate the urn of size n. Under both of these beliefs, the collapsed lottery corresponding to the smaller urn is the mean-preserving spread of that of the larger urn. For any belief formation with mean-preserving spread property, a DM exhibits preference for the larger urn if and only if she is ambiguity averse. Similarly, she is ambiguity neutral if and only if she does not have a strict preference on the size of ambiguous urns. Most of our ambiguity averse subjects, indeed, preferred the larger ambiguous urns to the smaller ones.

Similar arguments can be applied to the other two-stage models such as the Rank Dependent Utility Model of Yaari (1987) and Quiggin (1982). These models might be consistent with our experimental findings if we assume a concave utility function u, a convex probability transformation function, and assume that the decision maker's belief in the first stage of the uncertainty described above makes the smaller urn a mean preserving spread of the larger urn (see also Segal (1987) for this result).

Halevy and Feltkamp (2005) illustrate a case where a Bayesian decision maker would behave as if she is ambiguity averse in a two-stage process. The key assumption in this exercise is the fact that there are multiple draws from positively correlated urns. We can extend their intuition in our setup to make prediction on preference for urn size. If there are multiple draws from the same urn, such a DM will exhibit a strict preference for the larger urn.[14] However, as in our experiment, if there is only a single draw from an ambiguous urn, she is indifferent to betting on any size.

4.2. **Multi-Prior Models.** Multi-prior models describe a decision maker who holds multiple priors when the bet is ambiguous and evaluates the situation based on some aggregation of those priors. For example, the worst scenario in the multiple-prior set is used in the maxmin expected utility model of Gilboa and Schmeidler (1989). The models do not necessarily suggest any systematic relationship between the corresponding multi-prior sets of different ambiguous urns and hence can be thought of as consistent with any type of behavior in our setup. In order to make any prediction for a subject using multi-prior model, we need to establish how the multiple prior set varies with the urn size.

Our ambiguity averse subjects preferred larger urns over the smaller ones under ambiguity. Hence, in order to be in line with the maxmin expected utility model, the subjects' most pessimistic scenario on a smaller ambiguous urn should be more pessimistic than that on a larger ambiguous urn. An extremely pessimistic scenario may be more "plausible" on a smaller urn than on a larger one, i.e. the DM may think it must be impossible to have no beads with the color she bet on in a large urn. This intuition is in line with the model of Safra et al. (2018).

Ghirardato et al. (2004) introduced $\alpha-$maxmin expected utility model as a convex combination of the most pessimistic and most optimistic evaluations of an alternative with respect to a multiple prior set. When adopted to our experimental results, we may restrict this model by either imposing restrictions on the multiple prior set when the urn size changes or imposing restrictions on how the convex combination coefficient, α, changes with the urn size. If the most optimistic and pessimistic scenarios on a given ambiguous urn are independent of the urn size, then we need α to decrease with the urn size.[15] Alternatively, if the convex combination coefficient, α, remains fixed when the urn size changes, the worst and best scenarios of the multiple prior set should evolve to keep the average evaluation for a smaller urn worse than the average evaluation of a larger urn.

The Choquet expected utility model (see Gilboa (1987) and Schmeidler (1989)) uses capacity function v, a non-additive extension of probability measure. In our setup, the evaluation of an ambiguous urn with n beads will be $v(X_n)u(30)$ where X_n denotes the event that the color of the drawn bead from urn with n beads is X and the DM bets

[14]This conclusion is based on their assumption that the decision maker has uniform beliefs on each color combination.

[15]Note that ambiguity averse behavior of our subjects requires the convex combination coefficient of the $\alpha-$maxmin expected utility model to be greater than 0.5 if we also assume symmetric preference for betting on color Black and White in our experiments.

on color X on this urn. The general Choquet expected utility model is too flexible to make a prediction on the preference for the size of the urn, but our findings require that $0.5 > v(X_N) > v(X_n)$ where $n < N$.

4.3. **Source Models.** Abdellaoui et al. (2011) highlights the fact that probabilistic sophistication is usually violated when different sources of ambiguity are compared, even though it is satisfied within a single source. Tversky and Fox (1995) suggested that familiarity with the source generating uncertainty plays a role in the behavior of the DM. In our design, we use Ellsberg urns with no information on the distribution of colors for any urn size. This should minimize any asymmetric familiarity with one size urn compared to the other and keep the ambiguity amount based on information the same. To explain our data, the certainty equivalence calculation of an ambiguous bet on a given size urn needs to depend on the urn if a source model in the sense of Chew and Sagi (2008) is adopted.

5. Ambiguity in a Non-Experimental Setting: Investment Decisions

Below we provide some suggestive evidence from a non-experimental analysis to test an implication of our experimental findings in an application. Mutual fund portfolio choice shares significant similarities with the choice involving ambiguous urns in a typical Ellsberg experiment. First, investors have a similar information set regarding funds' assets, style, fees, and historical performance measures. More importantly, the lack of evidence on the persistence of performance in mutual fund returns suggests that investors face similar ambiguity regarding the future performance of the funds.[16]

Given these similarities, we argue that one can interpret the number of assets held by a mutual fund as corresponding to the number of beads in our experiment. If that is the case, we wonder whether investors take the number of assets held by a mutual fund (a measure of size) into account when deciding to allocate capital across different funds. In this non-experimental setting, we provide evidence consistent with the findings of our experiment. Controlling for commonly used determinants of fund flows (which include lagged raw and benchmark-adjusted returns, lagged fund flows, and total net assets as well as controls such as fund return volatility, fund fees, and fund age), we find that fund flows are positively and significantly related to the number of assets held by a fund. Having said that, as opposed to our experiment, data from real-world could be confounding. Hence, the evidence we offer in this section should be taken with a grain of salt.

[16]When testing the predictions of ambiguity aversion in a non-experimental setting, mutual funds have been a preferred choice of researchers. See, for example, Antoniou et al. (2015), Li et al. (2016).

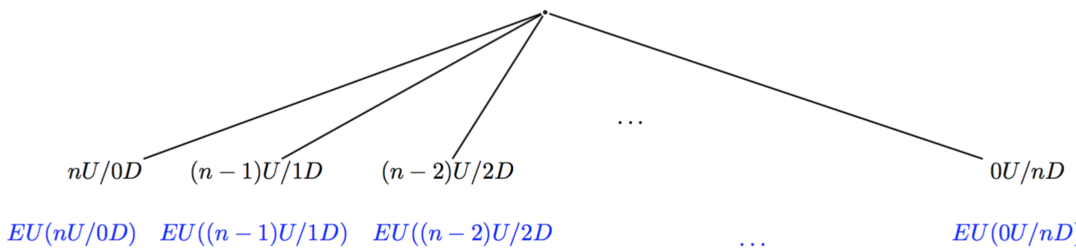


Figure 4. A Represetation of Mutual Fund

To facilitate comparison with the choice involving two ambiguous urns with a different number of beads, we measure fund size as the number of stocks held by a mutual fund.[17] Since the return of a fund is simply the weighted-average of the returns in the assets held by the fund, betting on a fund can be viewed as betting on an average asset in the fund. If we abstract away from other complications of portfolio choice problem, we can illustrate the similarity between betting on an urn with n beads in the experiment and investing in a mutual fund with n assets with equal weights. Let the price of each asset in the fund go up by $A\%$ or down by $B\%$, and the distribution of asset performances is unknown to the decision maker. If prices of all assets in a mutual fund increase, the investor earns $A\%$. Similarly, if all decrease, she loses $B\%$. Including other combinations of possible asset performances, there are $n+1$ payoff relevant states of the world where each state is characterized by the number of assets that increased or decreased. Since the investor views such a fund as an ambiguous option, we may represent this by an uncertainty illustrated in Figure 4, which is similar to Figure 3.

If investors prefer funds with a greater number of assets in their portfolios, we would then expect to observe a positive association between future fund flows (a measure of investment choice) and fund size, controlling for common determinants of fund flows used in the literature (lagged raw and benchmark adjusted returns, lagged fund flows, fund return volatility, fund fees, fund age, and total net assets of the fund).

Our empirical specification mainly follows Sirri and Tufano (1998). To estimate the impact of the number of assets in a fund's portfolio on investment decisions, we regress quarterly mutual fund flows on lagged values of various fund characteristics. We assume $flow_{i,t}$ is a linear function of the following variables and estimate a regression over 1981-2016:

$$(fundsize_{i,t-1}, flow_{i,t-1}, rawret_{i,t-1}, logTNA_{i,t-1}, exprat_{i,t-1}, volatility_{i,t-1})$$

[17] This measure of fund size is different from Total Net Assets (TNA) of the fund, which is a commonly used measure of size in mutual fund literature.

in which the dependent variable is the percentage flow to a fund the current quarter, t.[18] The independent variables, measured at the end of previous quarter *t-1*, include: (i) the number of holdings in a fund's portfolio ($fundsize$), (iii) lag fund flows ($flow$), (iii) average raw return over the previous year ($rawret$), (iv) log total net assets (TNA) of the fund, (v) fund return volatility in the previous year ($volatility$), and (vi) the expense ratio of the fund ($exprat$). Raw return is included to account for the flow-performance relation documented in the literature, while fund return volatility is included to account for preference for large size due to a diversification benefit, and the expense ratio is included to account for the documented negative association between fund fees and flows. We control for lag fund flows to account for persistence in quarterly fund flows. Finally, we control for TNA of the fund since previous literature documents a significant negative relation between the market value of a fund's assets and future fund performance (Berk and Green (2004), Chevalier and Ellison (1997), Ippolito (1992), Sirri and Tufano (1998)).

Table 6 reports the main findings.[19] The dependent variable in regressions (1) and (2) is quarterly fund flows. Specification 1 is our baseline flow regression, and it is consistent with the evidence documented in prior literature (e.g., Sirri and Tufano (1998) and Lou (2012)).[20] In specification 2, we include our variable of interest as an independent variable: *number of holdings in a fund's portfolio*. Consistent with the predictions from our experiment, we find that fund flows are positively related to the number of holdings in which a fund is invested. The coefficient of the *number of holdings* variable is both statistically significant (t-stat: 5.02) and economically meaningful. The magnitude of the coefficient implies that a one standard deviation increase in the number of holdings increases the

[18]Following prior literature (Sirri and Tufano (1998)), we compute quarterly fund flows as:

$$flow_{i,t} = \frac{TNA_{i,t} - (1 + r_t)TNA_{i,t-1}}{TNA_{i,t-1}}$$

in which TNA_t is a fund's total net assets at time t, and r_t is the fund's return over the prior quarter. Thus, the percentage fund flow is measured as the percent increase in the market value of the funds' assets in excess of the increase in market value due to performance of existing assets. Positive (negative) flow measure means that the fund experiences an inflow (outflow) of new investment.

[19]The sample in Table 6 is confined to domestic equity growth funds, as defined by either the fund's Lipper Objective Code, the fund's Strategic Insight Objective Code, or the fund's Wiesenberger Fund Type Code. All passive funds (i.e., fund names that contain any of the following words: index, idx, etf, russell, direxion, rydex, profund, wisdomtree s&p) and retirement funds (i.e., fund name contains any of the following years or words: 2005, 2010, 2015, 2020, 2025, 2030, 2035, 2040, 2050, 2055, 2060, or retirement) are removed. Funds that have a non-missing value for the ETF/ETN flag (et_flag) or the index flag ($index\_fund\_flag$) are removed. Funds that have "Y" for the variable annuity flag (vau_fund) are also removed. Funds that could not be identified either by name or by fund strategy are removed. Funds that have fewer than ten holdings at any given time are also removed. Missing expense ratios are backfilled up to eleven months, so that the expense ratio during that fiscal year is considered.

[20]Specifically, we find a significantly positive relationship between future fund flows and both lagged fund performance (*raw return*) and lagged fund flows. We also find a significantly negative relationship between fund flows and lagged fund volatility. The association between future fund flows and fund fees (*expense ratios*) is negative but insignificant.

subsequent quarter's fund flows by 0.25 percentage points. Given the average total assets under management of \$210.4 million in our sample, this amounts to an increase in the fund's assets by \$526,000 per quarter.

TABLE 6. Ambiguity Attitudes for the Fund Size (Number of Assets): Fund Flows and Performance

	Predicting future flows		***Predicting future returns***	
	(1)	(2)	(3)	(4)
Number of Holdings		0.00135***		0.00001
		(5.02)		(0.11)
Fund Flows (lagged)	0.314***	0.314***	0.00450**	0.00449**
	(18.08)	(18.05)	(2.38)	(2.38)
Average Raw Return	3.154***	3.153***	-0.0216	-0.0216
	(20.01)	(20.02)	(-0.33)	(-0.33)
Log TNA	-0.00650***	-0.00668***	-0.000166	-0.000168
	(-11.75)	(-12.03)	(-1.00)	(-1.01)
Expense Ratio	-0.942	-0.809	-3.596***	-3.595***
	(-0.71)	(-0.63)	(-6.58)	(-6.57)
Volatility	-0.304***	-0.304***	0.316***	0.316***
	(-3.80)	(-3.78)	(7.50)	(7.50)
Intercept	-0.0171	-0.0169	0.0111**	0.0111**
	(-1.17)	(-1.16)	(-2.13)	(-2.13)
Observations	36,261	36,261	36,261	36,261
Adjusted R-Squared	0.1869	0.1871	0.7594	0.7594

The dependent variable in regressions (1)-(2) is Quarterly Fund Flows, and the dependent variable in regressions (3)-(4) is Quarterly Returns. *Fund Flows* is quarterly fund flows and is winsorized by 1% at both the low and high end each month. Quarterly Returns is the quarterly raw return of the fund. *Average Raw Return* is the average monthly raw return of the fund over the previous twelve months. *Log TNA* is the natural logarithm of the total net assets at the end of the previous quarter. *Expense Ratio* is the monthly expense ratio of the fund during that fiscal year. *Volatility* is defined as the standard deviation of monthly returns over the previous twelve months. *Number of Holdings* is the total number of holdings in the fund at the end of the previous quarter (in hundreds). Quarter fixed effects are included in all regression specifications, and standard errors are clustered at the fund level. t-statistics are shown in parentheses, and statistical significance at the 1%, 5%, and 10% levels is denoted by ***, **, and *.

To test if mutual fund investors are ex-post rewarded by choosing a fund with a greater number of holdings, we re-estimate the regressions in (1) and (2) by using future quarterly returns as the dependent variable. Regressions (3) and (4) document the results.[21] More importantly, after controlling for common determinants of future fund performance, we do not find a significant association between the number of holdings in a fund and future

[21]For robustness, we re-estimate all the regression specifications by (i) using risk-adjusted returns instead of raw returns, (ii) controlling for fund flows up to the previous four quarters, and (iii) controlling for fund age. In all specifications, we obtain similar results. There is a significantly positive association between the number of assets in a fund and future fund flows and no relationship between the number of assets and future fund performance.

fund performance. These results provide evidence that investors prefer funds with a greater number of assets, but they do not seem to be rewarded for their preference in these funds.

6. Conclusion

In this paper, we experimentally show that individuals have a preference for larger ambiguous urns over smaller ones. In light of our experimental findings, we revisited the ambiguity models in terms of the restrictions our findings impose on them.

Ellsberg urn experiments have been the standard tool for demonstrating and estimating the ambiguity attitudes. However, no attention has been given to the size of the urn when comparing ambiguous urns. Our results highlights the importance of considering different urn sizes while estimating ambiguity attitudes. It is possible that the individuals may perceive larger urns as less ambiguous and behave as if they were less ambiguity averse; hence, those estimations will be biased.

Our experiments also contribute to the psychology literature on ratio bias. While we do find evidence of ratio bias in our data, subjects are heterogeneous, and the majority of subjects do not exhibit such bias with even ratios. To the best of our knowledge, ours is the first experiment to test ratio bias by eliciting strict preferences on *50-50* chances by using incentivized methods.

Finally, we analyze the mutual fund choices of investors by interpreting this investment problem as a choice of Ellsberg urn size. The data verifies that the investment is affected by the number of holdings in a fund, and there is a tendency to choose funds with more holdings even after controlling for commonly used determinants of fund flows. This means that the number of holdings in a fund should be controlled for in relevant regressions. This provides evidence for preference for size under ambiguity in a non-experimental setting.

References

Abdellaoui, M., Baillon, A., Placido, L., and Wakker, P. (2011). The rich domain of uncertainty: Source functions and their experimental implementation. *The American Economic Review*, 101(2):695–723.

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

Antoniou, C., Harris, R. D., and Zhang, R. (2015). Ambiguity aversion and stock market participation: An empirical analysis. *Journal of Banking and Finance*, 58:57–70.

Bade, S. (2015). Randomization devices and the elicitation of ambiguity-averse preferences. *Journal of Economic Theory*, 159:221–235.

Baillon, A., Halevy, Y., and Li, C. (2015). Experimental elicitation of ambiguity attitude using the random incentive system. Technical report, Vancouver School of Economics.

Barratt, A., Howard, K., Irwig, L., Salkeld, G., and Houssami, N. (2005). Model of outcomes of screening mammography: information to support informed choices. *British Medical Journal*, 330(7497).

Becker, G. M., DeGroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Systems Research and Behavioral Science*, 9(3):226–232.

Becker, S. W. and Brownson, F. O. (1964). What price ambiguity? or the role of ambiguity in decision-making. *Journal of Political Economy*, 72(1):62–73.

Berk, J. B. and Green, R. C. (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112(6):1269–1295.

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

Chevalier, J. and Ellison, G. (1997). Risk taking by mutual funds as a response to incentives. *Journal of Political Economy*, 105(6):1167–1200.

Chew, S. H., Miao, B., and Zhong, S. (2017). Partial ambiguity. *Econometrica*, 85(4):1239–1260.

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

Crockett, S., Izhakian, Y., and Jamison, J. C. (2019). Ellsberg's hidden paradox. Technical report, Baruch College Zicklin School of Business.

Denes-Raj, V. and Epstein, S. (1994). Conflict between intuitive and rational processing: when people behave against their better judgment. *Journal of Personality and Social Psychology*, 66(5):819–829.

Einhorn, H. J. and Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological review*, 92(4):433.

Einhorn, H. J. and Hogarth, R. M. (1986). Decision making under ambiguity. *Journal of Business*, pages 225–250.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, 75(4):643–669.

Epstein, L. G. (2010). A paradox for the "smooth ambiguity" model of preference. *Econometrica*, 78(6):2085–2099.

Epstein, L. G. and Halevy, Y. (2018). Ambiguous correlation. *Review of Economic Studies*, 86(2):686–693.

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

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

Gilboa, I. (1987). Expected utility with purely subjective non-additive probabilities. *Journal of Mathematical Economics*, 16(1):65–88.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153.

Gillen, B., Snowberg, E., and Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy*, 127(4):1826–1863.

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

Halevy, Y. and Feltkamp, V. (2005). A bayesian approach to uncertainty aversion. *The Review of Economic Studies*, 72(2):449–466.

Ippolito, R. A. (1992). Consumer reaction to measures of poor quality: Evidence from the mutual fund industry. *The Journal of Law and Economics*, 35(1):45–70.

Kirkpatrick, L. A. and Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: further evidence for two conceptual systems. *Journal of Personality and Social Psychology*, 63(4):534.

Klibanoff, P., Marinacci, M., and Mukerji, S. (2005). A smooth model of decision making under ambiguity. *Econometrica*, 73(6):1849–1892.

Klibanoff, P., Marinacci, M., and Mukerji, S. (2012). On the smooth ambiguity model: A reply. *Econometrica*, 80(3):1303–1321.

Klibanoff, P., Mukerji, S., and Seo, K. (2014). Perceived ambiguity and relevant measures. *Econometrica*, 82(5):1945–1978.

Lefebvre, M., Vieider, F. M., and Villeval, M. C. (2011). The ratio bias phenomenon: fact or artifact? *Theory and Decision*, 71(4):615–641.

Li, C. W., Tiwari, A., and Tong, L. (2016). Investment decisions under ambiguity: Evidence from mutual fund investor behavior. *Management Science*, 68(3):2509–2528.

Loomes, G., Starmer, C., and Sugden, R. (1989). Preference reversal: information-processing effect or rational non-transitive choice? *The Economic Journal*, 99(395):140–151.

Lou, D. (2012). A flow-based explanation for return predictability. *The Review of Financial Studies*, 25(12):3457–3489.

Maafi, H. (2011). Preference reversals under ambiguity. *Management science*, 57(11):2054–2066.

Machina, M. J. and Siniscalchi, M. (2014). Ambiguity and ambiguity aversion. In *Handbook of the Economics of Risk and Uncertainty*, volume 1, pages 729–807. Elsevier.

Neilson, W. S. (2010). A simplified axiomatic approach to ambiguity aversion. *Journal of Risk and Uncertainty*, 41(2):113–124.

Pinto-Prades, J.-L., Martinez-Perez, J.-E., and Abellán-Perpiñán, J.-M. (2006). The influence of the ratio bias phenomenon on the elicitation of health states utilities. *Judgment and Decision Making*, 1(2):118–133.

Pulford, B. D. and Colman, A. M. (2008). Size doesn't really matter: Ambiguity aversion in ellsberg urns with few balls. *Experimental psychology*, 55(1):31–37.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343.

Safra, Z., Segal, U., et al. (2018). A lot of ambiguity. Technical report, Boston College Department of Economics.

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

Segal, U. (1987). The ellsberg paradox and risk aversion: An anticipated utility approach. *International Economic Review*, 28(1):175–202.

Seo, K. (2009). Ambiguity and second-order belief. *Econometrica*, 77(5):1575–1605.

Sirri, E. R. and Tufano, P. (1998). Costly search and mutual fund flows. *The Journal of Finance*, 53(5):1589–1622.

Tversky, A. and Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological review*, 102(2):269.

Valenzuela Jr, A. (2003). Day labor work. *Annual Review of Sociology*, 29(1):307–333.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, 55(1):95–115.

Online Appendix

Instructions

Introduction:

Welcome to the experiment. In this experiment, you will make decisions on uncertain scenarios. The precise rules and procedures that govern the operation of these decisions will be explained to you below.

Various research foundations have provided funds for this research. The instructions are simple, and if you follow them carefully and pay attention to your decisions you can finish the experiment with a considerable amount of money, which will be paid to you in cash at the end. You will receive $7 participation fee for completing the experiment and some additional amount that will depend on the decisions you make during the experiment. The experiment will last about 1 hour. **Please do not talk to each other** during the session. If you have any question, please raise your hand and the experimenter will come and answer you.

Your task:

In this experiment, there are 7 choice problems with two versions of each: Version A and Version B. Hence you will make 14 choices in total. For each problem you are asked to make a choice between two options. Each option is a bet on the color of a bead that will be drawn randomly from an urn at the end of the experiment. The two options you compare in each problem will be about two urns that may have different size or composition of black and white beads. First you will choose a color (Black or White) that you want to bet on. Then in each problem, you will be asked to choose between the two urns specified in that problem.

Selecting the relevant decision problem for payment:

Before you make any choices, one of the choice problems will be selected at random according to the protocol specified in the following paragraph, and your chosen bet in that choice problem will determine your payment.

To select the choice problem that will determine your payment, the experiment coordinators will roll one 7-sided die that produces a number from 1 to 7 and one 6-sided die that produces a number from 1 to 6 for each participant. They will write the numbers on notes and put them into sealed envelopes that will be distributed to you. The numbers in your envelope will correspond to the choice problem that will determine your payment. The outcome of the 7-sided die determines the choice problem, and the outcome of the 6-sided die determines the version of the corresponding choice problem. If the 6-sided die's outcome is an odd number then version A of the corresponding choice problem will be used for your payment and if it is even then version B will be used for payment. Please write your name on the envelope and <u>do not open the envelope</u>. This protocol of determining payments is to make sure that you choose in each choice problem as if it is the question determining your payment.

[Determine the numbers, prepare the envelopes, and give them to the subjects]

Choosing Colors:

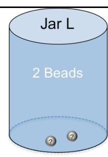
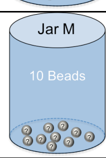
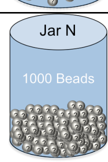
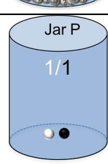
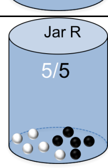
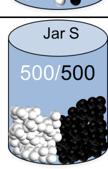
In the first part of the experiment, we will present you with six urns that are on the experimenter's desk. Each urn is filled with black and white beads. There is no other color besides black and white. We will tell you the number of beads in each urn but the composition of the two colors may or may not be told to you depending on the choice problem.

[Experimenter presents each Urn]

Next you will specify a color to bet on for each urn. <u>At the end of the experiment, one bead will be drawn from each urn.</u> If the color of the drawn bead matches with the color you specified for the urn you choose in your paying choice problem then you will receive a prize.

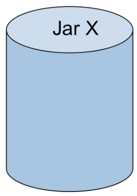
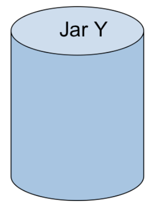
For example, let's say you specify Black for the Urn with two beads-unknown colors and this is the urn you choose in the choice problem selected for payment. If the bead drawn from this urn at the end of the experiment is also Black, then you will receive the specified prize for that choice problem. If the drawn bead is White, you will receive $0 for this choice problem.

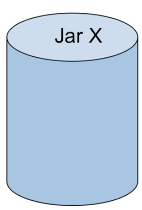
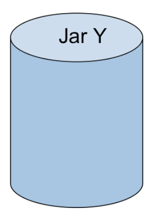
Now, please put a check mark (√) under the color that you want to bet on for each Urn in the table below:

URN	Black	White
		
		
		
		
		
		

Example of Binary Choice Problems:

Next you will be asked to choose between two urns in 7 choice problems where each problem will have version A and version B. An example of a choice problem is below:

Version A: *Please put a check mark (√) for the urn that you want to bet*	
 *If the color of the bead drawn from this urn matches with the color you specified for this urn initially then you will be paid* ***$30.25***	 *If the color of the bead drawn from this urn matches with the color you specified for this urn initially then you will be paid* ***$30***

Version B: *Please put a check mark (√) for the urn that you want to bet*	
 *If the color of the bead drawn from this urn matches with the color you specified for this urn initially then you will be paid* ***$30***	 *If the color of the bead drawn from this urn matches with the color you specified for this urn initially then you will be paid* ***$30.25***

At the end of the experiment, one bead will be drawn from each urn.

Let's say that you choose Urn Y in version A, your fixed color for this urn is White, and this is the problem that is written in your envelope. Then, at the end of the experiment if the drawn bead from Urn Y is indeed White, you will receive $30. If it is Black, you will receive zero (both are in addition to the payment of $7 you received for arriving to the experiment on time).

All the choice problems will be similar to the one in the Example. Note that version A and version B in this example are quite similar. Both A and B ask you to choose between Urn X and Urn Y. The prizes are $30 or $30.25. The difference is that Urn Y's prize is $30 in version A and it is $30.25 in version B (vice versa for Urn X). In all the problems you will answer in this experiment, you will choose between two urns with the prizes $30 or $30.25 and we will give you A and B versions of that choice problem.

You may choose any bet in any problem. There is no best decision that works for everyone. If you choose Urn Y that pays $30 in version A above, it means that you think your chance of winning in Urn Y is higher than your chance of winning in Urn X. Since, Urn Y pays 25 Cents more prize in version B, you should also choose Urn Y in this version.

You may choose $30 prize urn in one version of the problem and $30.25 prize urn in the other version. Similarly, you may choose $30.25 prize urn in both versions. However, choosing the urn that pays $30 in each version is not a good strategy as explained above.