COMPARING AMBIGUOUS URNS WITH DIFFERENT SIZES*

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

Abstract. We investigate experimentally preferences between different ambiguous processes generated by two-color Ellsberg urns. By providing symmetric information on urns with different numbers of beads and keeping the information on the most optimistic, pessimistic, and equal probability of winning 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. We analyze the role of ambiguity aversion and ratio bias of subjects in this behavior. We study the restrictions that our findings impose on the existing ambiguity models.

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). Ellsberg's thought experiment has been widely confirmed in numerous laboratory

*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, Bin Miao, 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. Yusufcan Masatlioglu gratefully acknowledgs financial support from the National Science Foundation through grant SES-1628883.

† University of Maryland, Department of Economics. E-mail: efozbay@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.

experiments (see e.g., Camerer and Weber, 1992 and Machina and Siniscalchi, 2014 for detailed surveys). In order to explain such behavior, normatively or prescriptively appealing theories of ambiguity have emerged (see Gilboa and Schmeidler, 1989, Schmeidler, 1989, Ergin and Gul, 2009, Klibanoff et al., 2005, Neilson, 2010, and Seo, 2009 and see also Machina and Siniscalchi, 2014 for a survey of ambiguity models). These models commonly predict preferences for betting on known distributions rather than unknown ones.

In this paper, we investigate whether the number of beads in the Ellsberg's two-color urn matters. Note that none of the well-known theories of ambiguity is crafted to make a prediction regarding the urn size. Consider a DM who decides to place a bet between two ambiguous Ellsberg's two-color urns, one with n beads and another with m beads with $m > n$, 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. Would the DM prefer the urn with m or n beads, or would she be indifferent between them?

Our laboratory experiments investigate size effect under ambiguity and its interaction with the ambiguity attitude and ratio bias. We elicit subjects' preferences between ambiguous urns with different sizes, risky urns with different sizes, and risky and ambiguous urns with the same size. See Figure 1 for the illustration of how this design elicits different types of size preferences. 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. Some subjects exhibited preferences for the urn size even when they compare two risky urns with different sizes but equal chances of winning on each; indicating ratio bias. We further study in detail the contribution of ambiguity aversion and ratio bias to the preferences for a larger urn under ambiguity.

Figure 1. Experimental Design

$$
\begin{tikzcd}
A2 \arrow[r, leftrightarrow, "\text{size effect}"] \arrow[d, leftrightarrow, "\text{ambiguity attitude}"] & A10 \arrow[r, leftrightarrow, "\text{size effect}"] \arrow[d, leftrightarrow, "\text{ambiguity attitude}"] & A1000 \arrow[d, leftrightarrow, "\text{ambiguity attitude}"] \\
R2 \arrow[r, leftrightarrow, "\text{ratio bias}"'] & R10 \arrow[r, leftrightarrow, "\text{ratio bias}"'] & R1000
\end{tikzcd}
$$

Notes: An and Rn denote betting on ambiguous and risky urns with n beads, respectively. The chance of winning is 50% on risky urns. Top horizontal arrows elicit size effect under ambiguity, the vertical arrows elicit ambiguity attitude, and the bottom horizontal arrows elicit size effect under risk (i.e., ratio bias).

The DM may be indifferent between betting on ambiguous urns with different sizes, as the provided information on each urn is the same, or, alternatively, she may prefer betting on one ambiguous urn rather than the other, as she may think that her chance of winning on one of them is higher. For example, say the colors of the beads in an urn are determined by throwing them into two adjacent black and white paints from a distance of 20 feet, as in the example of Einhorn and Hogarth (1985). 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 painted in the color that she had bet on among those 1000 beads. Hence, an ambiguity averse DM may exhibit a preference for a larger urn. Alternatively, the DM may dislike having a large number of possibilities in an uncertain situation (see Einhorn and Hogarth, 1986). While in a risky urn with two beads and 50% chance of winning, there is only one possibility of color composition (i.e., 1 black and 1 white beads), in an ambiguous urn with 2 beads there are 3 color composition possibilities. Similarly, in an ambiguous urn with 1000 beads, there are 1001 possibilities. It may be harder for the DM to contemplate 1001 possibilities than 3 possibilities, and hence, a larger urn may be perceived as less desirable. In that case, an ambiguity averse DM may exhibit a preference for a smaller urn. In this paper, we elicit subjects' preferences on Ellsberg's two-color urns with different number of beads. If the urn size is playing a role in the decision, this information, as a measure of the size of the ambiguous state space or complexity of the source of ambiguity, can be incorporated into the ambiguity models.

In many decision problems, individuals decide among different ambiguous situations where the size of the possible possibilities may have an effect on the perception 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 more workers gather at one retail location than another, would the size of the crowd (as the size of an Ellsberg urn) matter for the DM even if this information did not indicate anything about the chance of getting a better service? Alternatively, consider a situation where a prize is randomly assigned to a lucky winner, similar to Charlie's decision problem in the story of "Willy Wonka & the Chocolate Factory." Charlie wants to buy a Wonka chocolate bar because some of these bars include a Golden Ticket for a full tour of a mysterious chocolate factory as well as a lifetime supply of chocolate. Charlie may go either to a small store that carries few Wonka Bars or to a giant store that carries a lot of Wonka Bars. Which store would Charlie prefer to purchase his Wonka Bar? Our experiment explicitly addresses this type of questions in a context-free environment, focusing on the number of different compositions of winning/losing possibilities controlling for all other effects.

Based on our experimental findings, we revisit the existing models and discuss what our results impose on these models. We argue that multi-prior models and source models are too flexible and they can explain any behavior in our setup. On the other hand, two-stage models such as the smooth ambiguity model of Klibanoff et al. (2005) can accommodate our data on size effect under ambiguity and its interactions with ambiguity attitude.[1] Finally, we provide some calibrations based on the smooth ambiguity model for the size premium under ambiguity in order to quantify our findings.

[1]This framework is similar to Segal (1987) where ambiguous prospects are analyzed as lotteries with two stages, the Reduction of Compound Lotteries axiom is relaxed, and a rank dependent expected utility model (Quiggin, 1982 and Yaari, 1987) is applied at each stage by using the same utility function.

There are other ambiguity experiments that require subjects to compare different ambiguous processes (see Pulford and Colman, 2008, Halevy and Feltkamp, 2005, Abdellaoui et al., 2011, Epstein and Halevy, 2018, and Chew et al., 2017). Among those, Chew et al. (2017) is the most closely related design to ours as they also vary the possible compositions of the ambiguous urns and their decision problems complement ours. We discuss these experiments in detail in Section 5. Pulford and Colman (2008) repeated the standard Ellsberg experiment varying the urn sizes. Although our focus is to understand the preferences between two ambiguous urns with different sizes, we also ask the standard Ellsberg questions varying the urn 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 that they do not investigate-, we find that the percentage of subjects choosing the risky urn is significantly smaller. We discuss the differences between our design and theirs in more detail in Section 5.

Finally, we conduct an empirical investigation to assess the extent to which our evidence on the set of possible compositions of an ambiguous urn being a relevant component of decision exists in an application where ambiguity models are often used. Specifically, we investigate whether investors take the number of assets held by a mutual fund into account while investing. The resolution of uncertainty regarding a fund's performance is admittedly not identical to our abstract two-color Ellsberg urn in the experiment. However, in loosely speaking, one may argue that the number of assets held by a mutual fund may be considered as a measure for the size of the ambiguous state space in this environment. Controlling for the commonly used determinants of fund flows, we find that investors prefer mutual funds with more holdings even though this does not result in better investment performance.

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 summarizes other experiments that are closely related to ours. Section 6 discusses the relevance of the number of assets in a mutual fund in an investment decision based on an empirical analysis. Section 7 concludes. Additional analysis that is not included in the main text is presented in Appendix A, B, and C. The instructions for the experiment are presented in Appendix D.

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.[2] 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. 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 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.

At the beginning of a session, each subject signed an informed consent form and received the written instructions provided in Appendix D. 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 of eliciting ordinal rather than cardinal preferences to

[2]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.

avoid complications introduced by other mechanisms.[3] The subjects 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.[4] 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 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, we asked subjects to pick a color to bet on for each urn. The purpose of the selection was twofold: (i) to convince the subjects that the experimenters did not have any bias toward a particular color, and (ii) 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.

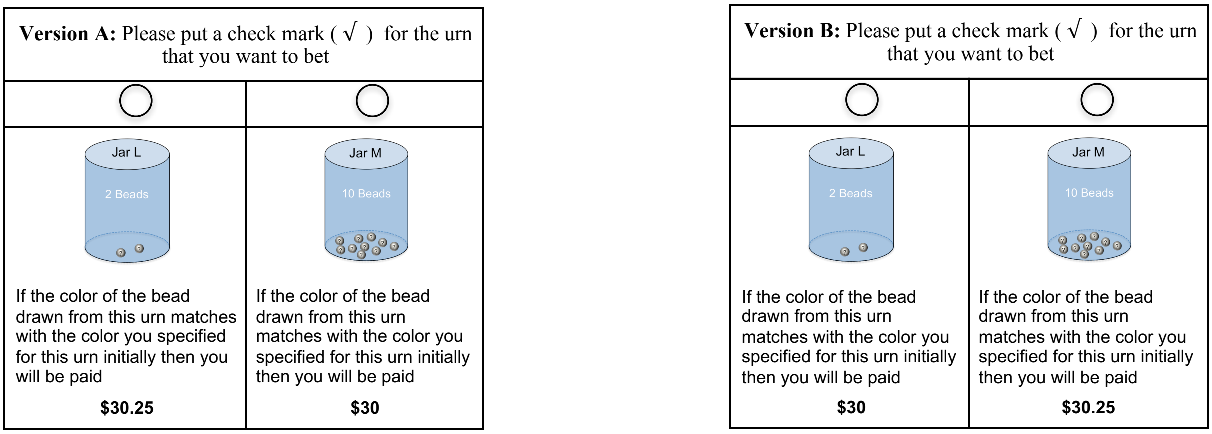
A typical alternative in a binary decision is a bet on an urn that pays a positive prize if the initially selected color of the subject for this urn matches the color of the randomly drawn bead from the urn at the end of the experiment. If the subject's selection of color and the experimenter's draw 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 2 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

[3]The cardinal valuations of the bets is typically elicited by mechanisms such as BDM mechanism of Becker et al. (1964), which is used by Halevy (2007), or a choice list method used by Abdellaoui et al. (2011) and Abdellaoui et al. (2015). These methods introduce other complexities that we wanted to avoid.

[4]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 resolution of this randomization 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).

Figure 2. Versions of a Sample Decision Problem



(a) Version A (b) Version B

maker between A2 and A10.[5] 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 A10 change in the 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. Even though observing a subject choosing the urn with a higher prize in both versions does not reveal anything about the preferences between urns, we believe that $0.25 was a small enough prize difference to elicit strict preferences of most subjects between urns with different sizes. Finally, if a subject picks the urn that pays $30 in both versions, she leaves money on the table.

[5]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.

Figure 1 in Section 1 summarizes the seven binary comparisons presented to the subjects. Remember that there were two versions of each comparison. The horizontal arrows on the top (A2 vs. A10 and A10 vs. A1000) are to elicit preferences for the size of ambiguous urns. The horizontal arrows at the bottom (R2 vs. R10 and R10 vs. R1000) elicit preferences for the size of risky urns (ratio bias). Finally, the three vertical arrows (A2 vs R2, A10 vs. R10, and A1000 vs. R1000) elicit ambiguity attitude varying the urn size. 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. 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 10 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, the researchers report 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 for choosing the smaller prize in both versions of the decision problem, and one subject was excluded for not answering all of the questions. Hence, the analyses are based on 116 subjects.[6] In our analysis, p-values are based on multinomial logistic regressions unless it is specified.

3.1. Size Preferences under Ambiguity and under Risk.

Size Preferences under Ambiguity and under Risk. We start with reporting the aggregate data on Table 1. 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 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 1), the majority of subjects strictly prefer the larger urn to the smaller urn (62.93% in A2 vs. A10 ($p = 0.000$) and 59.48% in A10 vs. A1000 ($p = 0.000$). 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.

[6]Choosing the smaller prize in both versions of the decision problems implies that the subject violates either monotonicity or transitivity. Our results do not change if we include these subjects.

A few subjects preferred the smaller urn to the larger one (8.62% in A2 vs. A10 and 10.34% in A10 vs. A1000).[7] Approximately one third of the subjects chose the higher prized urn under ambiguity. Table A1 in Appendix A reports the relation between the size preferences measured on A2 vs. A10 and A10 vs A1000.

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. R2 vs. R10 and R10 vs. R1000, and the percentage of subjects choosing the higher prize is greater than those choosing larger urn and smaller urn (for all such comparisons $p < 0.01$).[8] Table A2 in Appendix A reports the relation between the size preferences measured on R2 vs. R10 and R10 vs. R1000.

Table 1. 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.[9] According to our elicitation method, those subjects who

[7]The orders of presentation of the decision problems do not have a significant effect on the preference for the urn size under ambiguity (in A2 vs A10, $\chi^2(6) = 2.62, p = 0.855$; in A10 vs A1000, $\chi^2(6) = 2.52, p = 0.866$).

[8]The orders of presentation of the the decision problems do not have a significant effect on the preferences for the urn size under risk (in R2 vs R10, $\chi^2(6) = 7.45, p = 0.281$; in R10 vs R1000, $\chi^2(6) = 7.00, p = 0.320$).

[9]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 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

do not 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 1). While we observe less ratio bias than documented in the psychology literature[10], 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 preferences for the urn size under risk and ambiguity are observed at different degrees in the data. We will discuss the interaction between the two size effects in Subsection 3.3. Here, note that while no probability based theory can explain ratio bias, the ambiguity models are flexible enough to predict size effect under ambiguity. As noted by Figure 1, if the size effect is differentiated under risk and ambiguity, it will have implications for the ambiguity attitude when it is measured on a smaller versus a larger urn. Next we look at the relation between the urn size and ambiguity attitudes.

3.2. **Ambiguity Attitudes.** Table 2 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% on 2-, 10-, 1000-bead urns, respectively. Based on Wilcoxon matched-pairs signed-ranks tests, the percentages of subjects preferring the risky urn on 2-bead urns and 10-bead urns are not significantly different ($p = 0.819$). However, the percentage of subject preferring the risky urn on 1000 bead urns is significantly lower ($p < 0.05$).

subjects made sequential binary comparisons between urns. In their experiment, the small urn always offers 1−in−10 chance of winning a 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, and this might be an issue, as we discuss in our design.

[10] Denes-Raj and Epstein, 1994 report between 54% to 61% ratio bias when the winning chance is around 10%, and Dale and Rudski, 2007 report 41% ratio bias controlling for errors.

Table 2. 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%

In a typical experiment on ambiguity, 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.[11]

Importantly, existing experiments typically fix 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.[12] Although the percentages of the ambiguity averse subjects are not different in 2-bead and 10-bead urns,

[11]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 size preferences. 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 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.

[12]Most of our subjects either had a consistent measure of ambiguity attitude or switched attitude monotonically with the urn size. We had only 3 subjects whose ambiguity attitude was recorded as non-monotonic (e.g., averse on the smallest urn, seeking on the medium size urn, and averse again on the larger urn).

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 a 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. We cannot rule out that the ambiguity attitudes of the subjects stay the same across different urns, and the larger urns are perceived to be less ambiguous.

3.3. **Understanding the size effect at the individual level.** In order to see the correlations between ratio bias, ambiguity aversion, and the size preferences of subjects under ambiguity, we summarize the raw data based on the behavior in the three types of decision problems that the subjects answered. In this subsection, we do this exercise for decisions made on urns with 2 and 10 beads and repeat the similar analysis for urns with 1000 beads for robustness of the results in Appendix A. Note that in our three types of decision problems (size preferences under ambiguity –A2 vs A10–, size preferences under risk –R2 vs R10–, and ambiguity attitude –R2 vs A2–), 27 possible choices can be observed in the data. We report the frequencies of each type of behavior in Table A7 in Appendix A. Within this rich set of potential behaviors that can be observed, only five types of behavior are most frequent in our data, and the highest percentage among these was behaving ambiguity averse and preferring the larger urn under ambiguity while choosing the higher prize in the comparison of the risky urns.[14]

Table 3 presents the raw data to understand the correlation between the subjects' size preferences under ambiguity and their ambiguity attitudes. Note that among the 73 subjects, who preferred the larger urn under ambiguity (the third row of the table), 61

[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. However, they do not study when the size of the urn is 1000.

[14]Additionally, we see that only one subject violates transitivity on the overall choice problems, by choosing A10$\succ$A2$\succ$R2$\succ$R10$\succ$A10.

of them were ambiguity averse as well. This is the largest group of subjects in this table. This is the first evidence we have on the relationship between ambiguity aversion and preferences for larger ambiguous urn. We will quantify the effect of ambiguity attitude on subjects' size preferences further on Table 5.

Table 3. Size Preferences under Ambiguity and Ambiguity Attitudes

		R2 vs. A2 (R10 vs. A10)			
		Higher Prize	Risky Urn	Ambiguous Urn	Total
A2 vs. A10	Higher Prize	13 (12)	18 (17)	2 (4)	33
	Smaller Urn	0 (0)	7 (9)	3 (1)	10
	Larger Urn	5 (7)	61 (59)	7 (7)	73
	Total	18 (19)	86 (85)	12 (12)	116

Table 4 classifies the subjects based on their behavior when they compare two ambiguous urns and when they compare two risky urns. This is to understand the relation between size preferences under risk and ambiguity. Note that among those 73 subjects who preferred $A10$ over $A2$, most of them (39 subjects) did not exhibit ratio bias while 21 subjects also preferred larger urn under risk. Hence, the relationship between the preferences for betting on a larger urn under ambiguity and risk is weak. As we will see in the regressions of Table 5, the ratio bias is not a robustly significant determinant of preferences for larger ambiguous urns. Hence, we argue that size effect under ambiguity is not necessarily an amplification of size effect under risk.[15]

Tables 3 and 4 identify subjects based on their decisions on urns of sizes 2 and 10. The results are qualitatively the same when we repeat the analysis for the comparisons of urns of sizes 10 and 1000 (see Tables A4 and A5 in Appendix A.)

We also run OLS regressions to understand the differentiated effects of ratio bias and ambiguity aversion on size preferences under ambiguity. Table 5 reports these regression results where the choice of larger urn under ambiguity is regressed on a dummy for

[15]According to such amplification hypothesis, the biases should be amplified as the environment becomes more uncertain. For example, Maafi (2011) shows experimentally that the preference reversals under ambiguity are stronger than those under risk. In our data, 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.

TABLE 4. Size Preferences under Ambiguity and under Risk

		R2 vs. R10			
		Higher Prize	Smaller Urn	Larger Urn	Total
A2 vs. A10	Higher Prize	26	2	5	33
	Smaller Urn	4	4	2	10
	Larger Urn	39	13	21	73
	Total	69	19	28	116

ambiguity aversion, a dummy for ratio bias, and an interaction variable of these two dummies. The baseline is the behavior that is neither ambiguity averse nor exhibits preferences for larger urn under risk. For 2 vs. 10, the coefficient of ambiguity aversion dummy is significant at 1% level, while the ratio bias dummy is not significant. The ratio bias dummy becomes significant for the preferences between sizes 10 and 1000. This is expected, as subjects had stronger ratio bias on urn with 1000 beads and less ambiguity aversion when it is measured on larger urns. The interaction variable, corresponding to diff-in-diff, is not significant. The corresponding marginal effects of ambiguity aversion and ratio bias are different: fixing the distribution of ratio biased behavior, an ambiguity averse subject is about 71% likely to choose the larger ambiguous urn, while a non-ambiguity averse subject is only about 39% likely to choose it; fixing the distribution of ambiguity attitudes, having ratio bias increases the chance of choosing the larger ambiguous urn from 58% to 77%. Hence, the marginal effect of ambiguity aversion on size preferences is higher than the marginal effect of ratio bias.

Based on the results of Tables 4 and 5, we conclude that the subjects are heterogeneous in their aversion to ambiguity and ratio bias. While preferences for size under ambiguity is more pronounced for the subjects with ratio bias compared to subjects without ratio bias, the ratio bias cannot *solely* explain the size preferences under ambiguity. The ambiguity aversion of subjects plays a significant role in generating tendency to choose larger ambiguous urns. Moreover, as Table 5 indicates, ambiguity aversion always contributes to size effect under ambiguity but ratio bias does not. Hence, our subsequent theoretical discussion will focus on explaining the size effect within the ambiguity models.

Table 5. OLS Regressions of Preferences for Larger Ambiguous Urn

Dependent variable: Choosing the larger urn				
	2 vs. 10		10 vs. 1000	
	(1)	(2)	(1)	(2)
Ambiguity Aversion (AA)	0.338**	0.331**	0.518***	0.506***
	(0.116)	(0.121)	(0.104)	(0.105)
Ratio Bias (RB)	0.222	0.193	0.589**	0.607**
	(0.185)	(0.194)	(0.183)	(0.186)
AA∗RB	-0.052	-0.031	-0.370	-0.356
	(0.221)	(0.230)	(0.208)	(0.210)
Constant	0.333**	0.329**	0.125	0.117
	(0.101)	(0.118)	(0.087)	(0.106)
Order FE	No	Yes	No	Yes
N	116	116	116	116
Log-likelihood	-73.676	-73.404	-63.673	-62.15

Notes: Standard errors are in parentheses; $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

4. Incorporating the Size of the Urn into Ambiguity Models

In this section, we will discuss ambiguity models in conjunction with our empirical results. First, note that none of these probability based models is compatible with ratio bias. Hence, we focus on generating the preferences for size under ambiguity only in conjunction with ambiguity non-neutral behavior, consistent with our earlier results.

4.1. **Multi-Prior Models.**

Multi-prior models describe a decision maker who holds more than one prior 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.

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. Given the physical description of the urns in the experiment, there is no reason for the subjects to believe that the worst case color composition differs between urns.

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 either (i) by keeping the most optimistic and pessimistic scenarios on a given ambiguous urn independent of the urn size and requiring α to decrease with the urn size[16] or (ii) by fixing α and evolving the worst and best scenarios of the multiple prior set so that the average evaluation for a smaller urn becomes worse than the average evaluation of a larger urn when the urn size changes.

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 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 $n < N$.

4.2. **Source Models.** Tversky and Fox (1995) suggested that familiarity with the source generating uncertainty plays a role in the behavior of the DM.[17] In our design, we use Ellsberg urns with no information on the distribution of colors for any urn size minimizing any asymmetric familiarity with one size urn. If a source model in the sense of Chew and Sagi (2008) is adopted, the certainty equivalence calculation of an ambiguous bet on a given size urn needs to depend on the urn in order to explain our data.

4.3. **Two-Stage Models.** These models describe decision problems as two-stage lotteries where the uncertainty in the first stage is vaguely described, and the one in the second

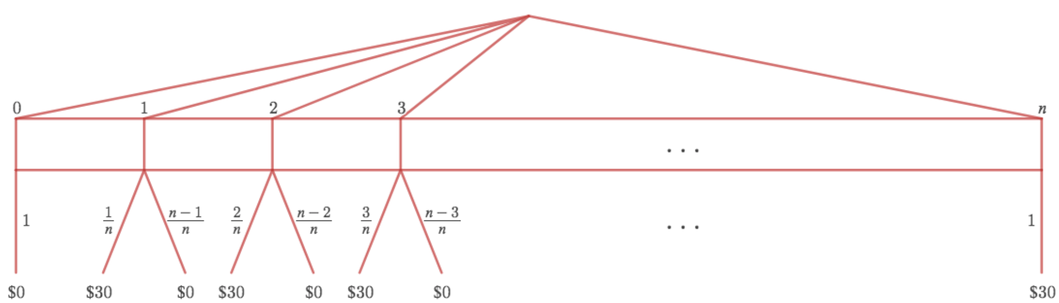
[16]Note that ambiguity averse behavior of our subjects requires the convex combination coefficient of the $\alpha-$maxmin expected utility model to be greater than 0.5 if we also assume symmetric preference for betting on color Black and White in our experiments.

[17]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.

stage is objectively specified. Motivated by the 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 (see Figure 3.) Note that the lotteries on the second stage have objective probabilities for the possible outcomes of zero and \$30. 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. 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 3. A Two-stage Compound Lottery Representation of Urn An



Notes: The 1st-stage represents the resolution of ambiguous states; the 2nd-stage represents the resolution of risky lottery. For each composition, the corresponding lottery on the 2nd-stage has objective probabilities for the possible outcomes of zero and \$30. The bar in the middle shows possible compositions of the ambiguous urn where the upper line represents the number of black beads.

The smooth ambiguity model of Klibanoff et al. (2005) in our setting evaluates the uncertainty described in Figure 3 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 the 2^{nd}-stage.

First, note that without any restriction on how p_i^n changes with n, any behavior can be rationalized by this model.[18]

Note that ambiguity aversion of a DM using the smooth ambiguity model is characterized by the concavity of ϕ. Also, recall that we found a strong connection between ambiguity aversion and size effect under ambiguity. Next we will show that concavity of ϕ can indeed explain the size effect.

Take the lottery described in Figure 3. If the induced distributions of expected utilities created from a larger urn is a mean preserving spread of the induced distribution of expected utilities created from a smaller urn, then having a concave ϕ function implies a preference for a larger ambiguous urn.[19] Recall that most of our ambiguity averse subjects, actually preferred the larger ambiguous urns to the smaller ones (see Tables 3 and A4.) We next describe some 1^{st}-stage beliefs that make the implicitly defined simple lottery corresponding to the larger urn, second order stochastically dominates the simple 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 single underlying distribution of colors to be used while generating the urns, i.e. $p_i^n =$

[18]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 compositions ($r_0^n = r_n^n = \frac{1}{2}$). For any non-linear ϕ, the preferences of a decision maker holding q^2 and q^{10} are complete opposite of the preferences of a decision maker holding q^2 and r^{10}.

[19]Note that the mean of such simple lottery is the mean of the expected utilities which are the outcomes of this induced lottery.

$\binom{n}{i}\lambda^i(1-\lambda)^{n-i}$, 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 1^{st}-stage beliefs, the implied simple lottery, in the sense described above, corresponding to the smaller urn induces a distribution of expected utilities that is a mean preserving spread of that implied by the larger urn. Similarly, a linear ϕ captures having no strict preferences on the size of ambiguous urns.

In another two-stage model, Segal (1987) applies the Rank Dependent Utility (RDU) model (Yaari, 1987 and Quiggin, 1982) in a recursive sense to explain ambiguity attitudes. Segal (1987) states sufficient conditions on the probability weighting function for RDU to capture ambiguity aversion on two color Ellsberg urn. However, those are not sufficient for increasing size preferences under ambiguity for urns with 2 or more beads and uniform first stage beliefs.[20] In other words, it is possible to create examples of convex probability weighting function implying ambiguity aversion while predicting decreasing, rather than increasing, size preferences under ambiguity.

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 preferences for urn size. If there are multiple draws from the same urn, such a DM would exhibit a strict preference for the larger urn.[21] However, as in our experiment, if there is only a single draw from an ambiguous urn, she would be indifferent to betting on any size.

[20]For 1 bead case, while the smooth ambiguity model can explain ambiguity aversion, RDU cannot. Assuming indifference between the two colors to bet on is equivalent to having a symmetric first stage belief around 50-50 distribution in RDU. Such symmetric beliefs would allow for only uniform distribution as the first stage belief if the urn has only one bead, but that would imply indifference between 50-50 risky urn and ambiguous urn with 1 bead, i.e. ambiguity neutrality. We thank an anonymous reviewer for pointing this out.

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

As noted earlier, there cannot be a probability based theory predicting ratio bias, because such theories would treat the ratios 1/2 and 5/10 the same. However, models that can allow for ratio bias might magnify the size effect under ambiguity.

Calibrations of size premium for the Smooth Ambiguity Model

In order to quantify preferences for the urn size under ambiguity, next we calibrate the smooth ambiguity model assuming uniform distribution for the 1^{st}-stage beliefs.[22,23] We assume a constant absolute ambiguity aversion model and a constant relative risk aversion model.[24]

In order to quantify the diminishing marginal effect of urn size, we define a size premium measure following the idea of ambiguity premium by Cubitt et al. (2018). Let $u(.)$ be the von Neumann utility of money, and $V(.)$ denote the utility of betting on an urn under the smooth ambiguity model. The size premium of moving from an ambiguous urn with n beads to an ambiguous urn with m beads is defined as the difference between the certainty equivalents of the two based on the assumed parameterization: Size Premium$(n-m) = u^{-1}(V(\text{A}m)) - u^{-1}(V(\text{A}n))$.

Table 6 reports the size premium for the ambiguous urns used in the experiments based on the smooth ambiguity model assuming uniform 1st-stage beliefs, $\phi = 0.079$ and $\rho = 0.066$. The size premium for switching from A2 to A10 is calibrated as almost four times higher than the size premium for switching from A10 to A1000. For the same parameters, we calculate the ambiguity premium of urn size 2, 10, and 1000 as about 4.7, 3.0, and 2.5, respectively, where ambiguity premium is defined as the difference between the certainty equivalents of risky and ambiguous urns of size n.[25]

---

[22] We chose this model for calibration purposes due to its precise parameterization that has been extensively studied by the literature.

[23] Uniform beliefs is reasonable with the physical environment of the experiments.

[24] Cubitt et al., 2018 also use these functional forms, $\phi(x) = -\dfrac{e^{-\phi x}}{\phi}$ and $u(x) = \dfrac{x^{1-\rho}}{1-\rho}$ in the estimations of the parameters. Note that $\phi > 0$ and $\phi < 0$ correspond to ambiguity-averse and ambiguity-seeking behaviors, respectively; $\rho = 0$, $\rho > 0$, and $\rho < 0$ correspond to risk-neutral, -averse, and -seeking behaviors, respectively.

[25] See Figure C1 in Appendix C for an illustration of certainty equivalents of ambiguous urns with different sizes.

Table 6. Size Premium under Smooth Ambiguity Model

Ambiguity Premium			Size Premium	
$A2$	$A10$	$A1000$	$(10-2)$	$(1000-10)$
4.721	2.958	2.489	1.763	0.468

Notes: The calculations of the table use the smooth ambiguity model with constant absolute ambiguity aversion with parameter $\phi = 0.079$ and constant relative risk aversion with parameter $\rho = 0.066$ for the 1^{st}-stage and 2^{nd}-stage functions and uniform 1^{st}-stage beliefs.

In our design, a subject is classified as someone who strictly prefers the larger urn under ambiguity if she prefers the larger urn both when its prize is \$30 and when it is \$30.25. In other words, the choice of A10(prize=30)$\succsim$ A2(prize=30.25) and A10(prize=30.25)$\succsim$ A2(prize=30) is identified as strict preference for the 10-bead urn over the 2-bead one. Similarly, a strict preference for the 1000-bead urn over the 10-bead one is identified if A1000(prize=30)$\succsim$ A10(prize=30.25) and A1000(prize=30.25)$\succ$A10(prize=30). One can calculate the minimum ambiguity attitude parameter ϕ that is needed to satisfy the inequalities above for each comparison. When we do that calculation assuming uniform 1st-stage beliefs and CRRA parameter of $\rho = 0.066$, we find the cutoff of $\phi = 0.004$ for $A10 \succ A2$ and the cutoff of $\phi = 0.017$ for $A1000 \succ A10$. This means that, according to this calibration, anyone who has an ambiguity aversion parameter larger than 0.017 will always prefer the larger urn (both in 2 vs 10 comparison and 10 vs. 1000 comparison). However, the ones whose $\phi \in (0.004, 0.017)$ will choose $A10$ over $A2$ but choose the higher prize when they compare $A1000$ and $A10$. Hence, more subjects are expected to strictly prefer $A10 \succ A2$ than $A1000 \succ A10$ as long as there are some subjects whose parameter $\phi \in (0.004, 0.017)$. This implication of the smooth ambiguity is consistent with our data. Recall that we have higher percentage of subjects who preferred the larger urn in the comparison of $A2$ vs. $A10$ than those who preferred the larger urn in the comparison of $A10$ vs. $A1000$ (62.93% vs. 59.48% in Table 1.) Since the drop in the percentages are relatively small in the data, this means that there were very few subjects whose ambiguity attitude parameter was in the interval of $(0.004, 0.017)$ if one takes the calibration exercise under the smooth ambiguity model as the basis.[26]

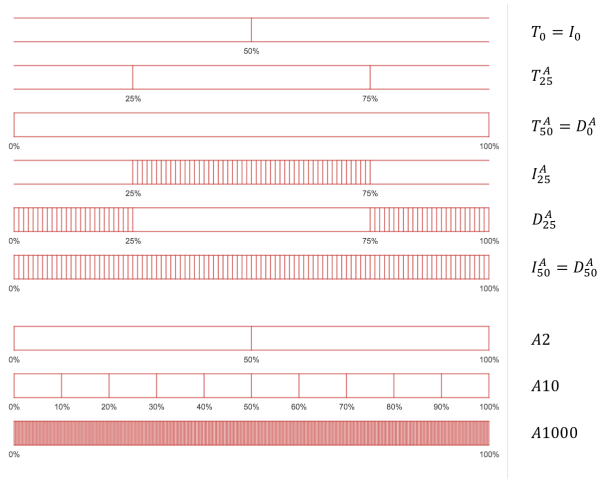
5. Related Experimental Literature

Chew et al. (2017) is the most closely related paper to ours where the authors extend the full ambiguity in two-color Ellsberg experiment to investigate the attitudes toward three distinct kinds of symmetric partial ambiguity. They create ambiguous bets in a deck of 100 cards with two different colors (red or black). In the interval ambiguity, denoted by I_n^A, a deck contains at least $50 - n$ and at most $50 + n$ of the same color cards. When the number of red or black cards is either not more than n or not less than $100 - n$, it is called disjoint ambiguity, denoted by D_n^A. Finally, when the number of red cards is either $50 - n$ or $50 + n$ with the rest of the cards are being black is called two-point ambiguity, denoted by T_n^A. They consider the cases of $n = 0$, 25, and 50. Figure 4 illustrates the possible chances of winning when betting on a certain color for the options used by Chew et al. (2017) and by our experiments. The bars on the first six rows correspond to the ambiguous bets used in Chew et al. (2017) and the last three correspond to our ambiguous urns of A2, A10 and A1000.[27] This joint illustration of two designs should clarify the comparison between the two experiments. Again, each vertical line in a row of Figure 4 indicates a possible color composition (i.e., possible chance of winning) for the corresponding bet. For example, since there are three possible color compositions for A2, possible chances of winning are 0%, 50%, and 100%. In the seventh row of Figure 4, these chances are denoted by the vertical line at the very left for 0% chance of winning, the vertical line in the center for the 50% chance of winning, and the vertical line at the very right for 100% chance of winning for someone who bet on color black.

[26]Since the parameter interval of $(0.004, 0.017)$ is on the lower tail of ambiguity aversion parameter distribution and much lower than the estimated average of 0.079 in Cubitt et al. (2018), one should not expect a large subject population with parameters in that interval.

[27]The first row corresponds to the known deck of 100 cards with 50 red cards and 50 black cards and denoted by T_0 or I_0 by Chew et al., 2017. The probability of winning for this bet is 50%.

Figure 4. Illustration of Bets Used in Chew et al. (2017) and in This Paper



Notes: The different forms of partial ambiguity used in Chew et al. (2017) are illustrated on the first six rows and ambiguous urns used in our experiments are illustrated on the last three rows. Each vertical line in a row indicates a specific chance of winning (i.e., possible urn composition) in the corresponding urn.

What is common in both designs is the fact that the possible chances of winning for each ambiguous bet of interest differ and both designs aim to understand how decision makers react to that. Chew et al. (2017) gradually add new possibilities which are either worse or better than the existing ones when they move from I_{25}^A to I_{50}^A (lines 4 and 6 in the figure).[28] In our design, the possible chances of winning of a smaller urn is always a subset of a larger urn, and 0%, 50%, and 100% chances of winning are always in the support.[29] While keeping this structure, we gradually add more possibilities in a uniform way when we move from a smaller urn to a larger one (see the last three lines of Figure 4.) We have the same worst (0%) and best (100%) possible chance of winning for our possible

[28]In the disjoint case, they gradually take away possible compositions from the middle range when the ambiguity moves from D_{50}^A to D_{25}^A and from D_{25}^A to D_0^A while keeping the worst and best possible compositions untouched.

[29]This is crucial for our design for different reasons: (i) 0% and 100% keep the extreme best and worst case scenarios constant across bets, (ii) 50% makes it possible to compare the ambiguous urns with the risky counterparts, (iii) 50% also eliminates any behavioral effects caused by having only extreme beliefs such as 0% and 100%. This design features are satisfied by only one of the bets in Chew et al. (2017) (the sixth row).

compositions, and add some new possibilities between the existing ones in a uniform sense when we move from a smaller ambiguous urn to a larger one. So, in terms of the types of new possible chances of winning added, while they eliminate chunks of possible chances of winning, we eliminate some possible compositions in a uniform sense. Hence, the two design complement each other.

The number of possible color compositions in the Ellsberg's urn might be a measure of degree of ambiguity (Becker and Brownson, 1964). Note that this absolute measure varies in both our and Chew et al. (2017) designs. One may also think a relative measure based on the number of possible compositions described to the subjects with respect to all the possibilities. In their design the absolute and relative measures move together. However, the relative measure is constant in our design while the absolute measure increases with the urn size. They find in a reduced form analysis that the subjects are averse to increasing degree of ambiguity which might be attributed to an increase in either absolute or relative measures in their design. Since our subjects preferred the larger ambiguous urns (i.e. higher absolute measure), together with Chew et al. (2017) finding, one may argue that the subjects are sensitive to the relative changes in the degree of ambiguity rather than the absolute one. Having said that, the two studies are consistent in their finding of aversion to spread (under the uniform distribution assumption for the belief on possible color compositions).

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.

In a typical ambiguity experiment, the urn size is fixed and subjects are asked to compare risky and ambiguous urns. Pulford and Colman (2008) repeat the standard Ellsberg experiment varying the urn sizes. They find that the percentages of subjects choosing the risky urn are not significantly different for the urns containing 2, 10, or 100 beads and conclude that the ambiguity attitude is not affected by the urn size.[30] Although our focus is to understand the preference between two ambiguous urns with different sizes, we also ask the standard Ellsberg questions by varying the size of the pair of risky and ambiguous urns. 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. Aside from that, we argue that due to our main findings of "preferences for larger urn under ambiguity," it may be misleading to make inferences about the robustness of ambiguity attitudes by 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.

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

When testing the predictions of ambiguity aversion in a non-experimental setting, mutual funds have been a preferred choice of researchers.[31] The lack of evidence on the persistence of performance in mutual fund returns suggests that investors face ambiguity regarding the future performance of the funds. In this section, we provide a suggestive evidence from a non-experimental analysis of mutual funds and test whether investors prefer mutual funds with larger fund size, as proxied by the number of assets in a fund's portfolio.

The resolution of uncertainty regarding a fund's performance is admittedly not identical to the abstract two-color Ellsberg urn in our experiment. However, one may argue that the number of assets held by a mutual fund may be considered as a measure for the size

[30] One should be careful interpreting the evidence in Pulford and Colman (2008) due to the way they generate ambiguous bets and elicit preferences. First, they interpret the choice of one alternative over another as a strict preference although indifference would imply the same behavior in that design. Second, they elicit the preferences between a compound lottery with known probabilities (which they interpret as an ambiguous urn) and its reduced form version (which they interpret as a risky urn). While the literature find high correlation between ambiguity neutrality and reduction of compound lotteries (see e.g. Chew et al., 2017; Dean and Ortoleva, 2019; Gillen et al., 2019; Halevy, 2007 with an exception of Abdellaoui et al., 2015), telling the subjects the probabilities is not the well-accepted methodology for generating ambiguity.

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

of the ambiguous state space in this environment, because it might be a proxy for the set of possible resolutions of ambiguous stage. Since we find the set of possible compositions of an ambiguous urn being a relevant component of decision in the experiments, we investigate whether such effect exists in the evaluation of mutual funds. 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. Appendix B describes the empirical methodology and presents the regression results in Table B1. To summarize, consistent with our conjecture, our empirical tests provide evidence that investors prefer mutual funds with a greater number of assets. Interestingly, these investors do not seem to be rewarded for their preference, as these funds yield lower returns in the future.[32]

Having said that, as opposed to our experiment, data from real-world could be confounding. The analogy between mutual fund size effect and size preferences identified in our experiment is subject to several underlying assumptions (i.e., homogeneous assets, correlation structure). Naturally, since a mutual fund holds multiple assets, a parallel experimental setting would involve bets on multiple Ellsberg urns (similar to Epstein and Halevy, 2018), which is beyond the scope of our paper. Therefore, the evidence we offer in this section should be taken with a grain of salt. Nevertheless, we hope that this empirical exercise helps encourage more empirical research in the area.

7. 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

[32]These results are also consistent with the preference for "over-diversification" explanation. However, since this section is intended to provide suggestive evidence for the empirical relevance of our results, we leave the more detailed analysis in this empirical setting to future work.

comparing ambiguous urns. Our results highlight 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.

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 Risk 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.

Cubitt, R., Kuilen, G., and Mukerji, S. (2018). The strength of sensitivity to ambiguity. *Theory and Decision*, 85(3):275–302.

Dale, D. and Rudski, J. (2007). Innumeracy and incentives: A ratio bias experiment. *Judgment and Decision Making*, 2(4):243.

Dean, M. and Ortoleva, P. (2019). The empirical relationship between nonstandard economic behaviors. *Proceedings of the National Academy of Sciences*, 116(33):16262.

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. 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., 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.

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.

Appendix A. Additional Analysis

Table A1. Size Preferences under Ambiguity by Urn Size

		$A10$ vs. $A1000$			
		Higher Prize	Smaller Urn	Larger Urn	Total
$A2$ vs. $A10$	Higher Prize	23	1	9	33
	Smaller Urn	2	2	6	10
	Larger Urn	10	9	54	73
	Total	35	12	69	116

Table A2. Size Preferences under Risk by Urn Size

		$R10$ vs. $R1000$			
		Higher Prize	Smaller Urn	Larger Urn	Total
$R2$ vs. $R10$	Higher Prize	56	1	12	69
	Smaller Urn	3	10	6	19
	Larger Urn	6	4	18	28
	Total	65	15	36	116

Table A3. Ambiguity Aversion and Ratio Bias for Subjects who Prefer Larger Ambiguous Urn

		Ratio Bias		
		No	Yes	Total
Ambiguity Averse	No	9	5	14
	Yes	43	16	59
	Total	52	21	73

Notes: Table restricts sample to those who exhibit preferences for larger urn under ambiguity, i.e., A10 $\succ$ A2. Ratio Bias is defined by R10 $\succ$ R2; Ambiguity Aversion is defined by R10 $\succ$ A10.

Table A4. Size Preferences under Ambiguity and Ambiguity Attitudes

		$R10$ vs. $A10$			
		Higher Prize	Risky Urn	Ambiguous Urn	Total
$A10$ vs. $A1000$	Higher Prize	13	17	5	35
	Smaller Urn	2	7	3	12
	Larger Urn	4	61	4	69
	Total	19	85	12	116

Table A5. Size Preferences under Ambiguity and under Risk

		$R10$ vs. $R1000$			
		Higher Prize	Smaller Urn	Larger Urn	Total
$A10$ vs. $A1000$	Higher Prize	32	1	2	35
	Smaller Urn	3	5	4	12
	Larger Urn	30	9	30	69
	Total	65	15	36	116

Table A6. Ambiguity Aversion and Ratio Bias for Subjects who Prefer Larger Ambiguous Urn

		Ratio Bias		
		No	Yes	Total
Ambiguity	No	3	5	8
Averse	Yes	36	25	61
	Total	39	30	69

Notes: Table restricts sample to those who exhibit preferences for larger urn under ambiguity, i.e., A1000 $\succ$ A10. Ratio Bias is defined by R1000 $\succ$ R10; Ambiguity Aversion is defined by R10 $\succ$ A10.

TABLE A7. Individual Types: Raw Data

Types under $n < m$				2 − 10		10 − 1000	
A_n vs A_m	R_n vs A_n	R_n vs R_m		Frequency	%	Frequency	%
0	0	0		11	9	12	10
0	0	1		0	0	0	0
0	0	2		2	2	1	1
0	1	0		14	12	16	14
0	1	1		2	2	1	1
0	1	2		2	2	0	0
0	2	0		1	1	4	3
0	2	1		0	0	0	0
0	2	2		1	1	1	1
1	0	0		0	0	1	1
1	0	1		0	0	1	1
1	0	2		0	0	0	0
1	1	0		3	3	1	1
1	1	1		3	3	2	2
1	1	2		1	1	4	3
1	2	0		1	1	1	1
1	2	1		1	1	2	2
1	2	2		1	1	0	0
2	0	0		0	0	2	2
2	0	1		2	2	0	0
2	0	2		3	3	2	2
2	1	0		35	30	27	23
2	1	1		10	9	9	8
2	1	2		16	14	25	22
2	2	0		4	3	1	1
2	2	1		1	1	0	0
2	2	2		2	2	3	3
			Total	116	100	116	100

Notes: Table codes each subject's choices between urn sizes 2 and 10 and urn sizes 10 and 1000 according to the subject's preferences for size under ambiguity (first column), ambiguity attitudes (second column), and preferences for size under risk (third column). In first three columns, "0" represents the behavior of choosing the urn with a higher prize in both versions. "1" represents the behavior of choosing the former option in the column header in both versions, and "2" represents the behavior of choosing the latter option in the column header in both versions. The fourth and sixth columns show the number of subjects behaving in the way corresponding row specifies in comparisons of sizes 2 versus 10, and 10 versus 1000, respectively. Columns five and seven report the same information in terms of percentage of subjects behaving in that way. The highlighted rows are the top five most frequent behavior in the data.

Appendix B. Mutual fund flows and the number of assets in fund's portfolio

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:

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

in which the dependent variable is the percentage flow to a fund in the current quarter, t.[33] 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 a 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 (Ippolito, 1992; Chevalier and Ellison, 1997; Sirri and Tufano, 1998; Berk and Green, 2004).

[33]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.

Table B1 reports the main findings.[34] 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).[35] In specification 2, we include our variable of interest as an independent variable: *number of holdings in a fund's portfolio*. 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 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.

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.[36] 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 fund performance.

[34]The sample in Table B1 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.

[35]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.

[36]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.

Table B1. 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

Notes: 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 *.

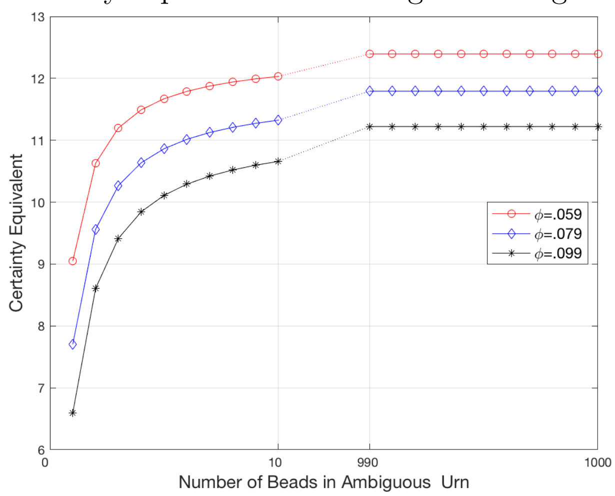
Appendix C. Calibrations of size premium for the Smooth Ambiguity Model

Figure C1 draws the certainty equivalent of betting on an ambiguous urn under smooth ambiguity as a function of the number of beads in the urn for ambiguity aversion parameters of $\phi = 0.059, 0.079,$ and 0.099. The figure is drawn for $\rho = 0.066$ which is the estimation of risk aversion parameter by Cubitt et al., 2018. They estimated $\phi = 0.079$.

We also draw the certainty equivalent functions for a little more and a little less ambiguity averse DM to illustrate the robustness of the size preferences under this model.

Note that for all three selections of the ambiguity parameter, ϕ, in Figure C1, the certainty equivalents are increasing with the urn size, which is expected as smooth ambiguity model predicts increasing size preferences under ambiguity aversion. The figure also shows that the certainty equivalent functions are concave in size. This means that the same increase in size will have a diminishing marginal effect on the preferences for size.

Figure C1. Certainty Equivalent of Betting on Ambiguous Urn of Size n



Notes: The figure illustrates the certainty equivalent of urn An with prize of \$30 under smooth ambiguity model with uniform 1$^{\text{st}}$-stage beliefs, a 1$^{\text{st}}$-stage function with constant absolute ambiguity aversion with parameter $\phi \in \{0.059, 0.079, 0.099\}$, and a 2$^{\text{nd}}$-stage function with constant relative risk aversion with parameter $\rho = 0.066$.

Appendix D. 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 do not open the envelope. 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.

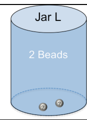
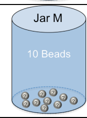
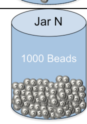
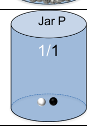
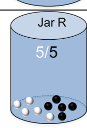
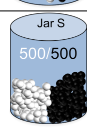
[Determine the numbers, prepare the envelopes, and give them to the subjects]

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]

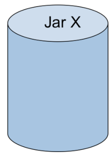
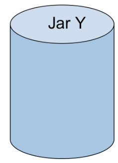
Next you will specify a color to bet on for each urn. At the end of the experiment, one bead will be drawn from each urn. If the color of the drawn bead matches with the color you specified for the urn you choose in your paying choice problem then you will receive a prize.

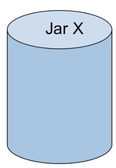
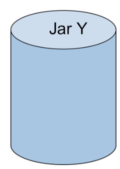
For example, let's say you specify Black for the Urn with two beads-unknown colors and this is the urn you choose in the choice problem selected for payment. If the bead drawn from this urn at the end of the experiment is also Black, then you will receive the specified prize for that choice problem.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