CESifo WORKING PAPERS

10745
2023

November 2023

A Two-Ball Ellsberg Paradox

Brian Jabarian, Simon Lazarus

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10745*

A Two-Ball Ellsberg Paradox

Abstract

We conduct an incentivized experiment on a nationally representative US sample (N=708) to test whether people prefer to avoid ambiguity even when it means choosing dominated options. In contrast to the literature, we find that 55% of subjects prefer a risky act to an ambiguous act that always provides a larger probability of winning. Our experimental design shows that such a preference is not mainly due to a lack of understanding. We conclude that subjects avoid ambiguity per se rather than avoiding ambiguity because it may yield a worse outcome. Such behavior cannot be reconciled with existing models of ambiguity aversion in a straightforward manner.

Keywords: uncertainty, complexity, ambiguity, decision-making.

Brian Jabarian
Booth Business School
University of Chicago / IL / USA
brian.jabarian@chicagobooth.edu

Simon Lazarus
Department of Economics
Princeton University / NJ / USA
slazarus1@hotmail.com

October 30, 2023
We are indebted to Leeat Yariv and Jean-Marc Tallon for their continued guidance. We are grateful for the comments of Miguel Ballester, Roland Bénabou, Gary Charness, Douglas Bernheim, Elias Bouacida, Ernst Fehr, Marc Fleurbaey, Evan Friedman, Yoram Halevy, Nicolas Jacquemet, Christoph Kuzmics, Yves Le Yaouanq, Pietro Ortoleva, Franz Ostrizek, Ryan Opra, Pëllumb Reshidi, Evgenii Safonov, Elia Sartori, Denis Shishkin, Richard Thaler, and our fellow Ph.D. colleagues at the Princeton Department of Economics for their generous participation into our pilots. We are grateful for the research assistance provided by Christian Kontz and Andras Molnar. We thank seminar audiences at CESifo, Gate-Lab, Chicago Booth, Princeton, MIT Sloan, and PSE. We thank the Princeton Experimental Laboratory for the Social Sciences for financial support. Princeton IRB approval # 12380 was obtained on April 29, 2020. This paper was partly written while Brian was visiting the Department of Economics at Princeton University and the Kahneman-Treisman Center for Behavioral Science and Public Policy in 2018-2020. He thanks their hospitality. He also acknowledges financial support from the Paris School of Economics, Sorbonne Economics Center, the Forethought Foundation, the Grant ANR-17-CE26-0003, and the Grant ANR-17-EURE-001 for his research visit to Princeton University.

1 Introduction

Are people willing to give up potential gains to avoid ambiguous situations, even when ambiguity can only benefit them? Since Ellsberg's famous paradox (Ellsberg (1961)), experiments have shown that people often exhibit ambiguity aversion because they fear that ambiguous situations may yield worse outcomes. In other words, people avoid situations where they cannot assign exact probabilities to possible outcomes, even if it means possibly giving up higher payoffs for fear of uncertainty resolving in a worse outcome. Scholars have developed models to accommodate such behavior. Such models include, among others, Choquet expected utility from Schmeidler (1989), Maximin expected utility from Gilboa and Schmeidler (1989), alpha-maximin expected utility from Ghirardato et al. (2004), as well as other proposals by Klibanoff et al. (2005), Maccheroni et al. (2006), and Strzalecki (2011).

By contrast with these previous works, this paper shows, in a simple incentivized experiment, that people frequently avoid ambiguity even when it can only result in better outcomes. Its design allows concluding that such behavior is neither entirely due to misunderstanding nor holding incorrect beliefs about the ambiguous situation. This result suggests that subjects have an inherent dislike for ambiguity, which is inconsistent with these models.

At the heart of our experiment is a "Two-Ball" gamble from Jabarian (2019): we have two urns, each containing red and blue balls. One is a risky urn with 50 red and 50 blue balls; the other is an ambiguous urn with unknown proportions of red and blue balls, as in Ellsberg's original thought experiment. The difference is that now, subjects draw two balls with replacement from one of these urns. Subjects win \$3 if the two balls have the same color. Would you rather play this gamble with the risky or ambiguous urn?

Independently of the color chosen, drawing from the risky urn gives a 50% chance of winning, while drawing from the ambiguous urn guarantees at least a 50% chance of winning, regardless of the proportions of red and blue balls. For example, if the ambiguous urn contains 60 red and 40 blue balls, its win probability is $.6^2 + .4^2 = .52$. This characteristic of the gamble entails that nearly all existing models require a decision-maker to choose the ambiguous urn over the risky urn. Despite this, 45% of the subjects in our experiment prefer the risky urn. Subjects were willing to pay 8.5% more for the risky gamble than the ambiguous one. We call this result the Two-Ball Ellsberg Paradox.

Unless subjects have beliefs over the two draws that are not consistent with the information given to them (say, subjects somehow believe the draws are not independent), the choice of RR over AA cannot be reconciled with existing models of ambiguity aversion straightforwardly. For instance, in the Maxmin Expected utility model of Gilboa and Schmeidler (1989), if subjects entertained the entire simplex over $\{R, B\}$ for the composition of an urn U and form beliefs over the two draws by composing each prior with itself, this would lead to indifference between gambles RR and AA. As in Gajdos et al. (2008), Dominance implies that RR cannot be strictly preferred to AA. If one assumes that the set of priors is a subset of $\{q \in \Delta(\{R, B\}^2) | q = p \times p; p \in \Delta(\{R, B\})\}$, then the choice RR over AA is incompatible with α-maxmin expected utility for any α. This choice is also incompatible with Savage (1954)'s Subjective Expected Utility model if beliefs are a product measure of the type $p \times p$.

In exposing the paradox, it is expedient to illustrate it as a choice between two distinct gambles. Nonetheless, we ascertain each gamble's certainty equivalent (CE) within our experimental setting by employing a multiple price list (MPL). We subsequently make a comparative analysis to discern whether subjects exhibit paradoxical behavior. This methodological decision is underpinned by several reasons. The survey by Jabarian (2021), conducted on a representative sample from the U.S., utilized a choice setup, offering an initial empirical indication supporting the paradox. The primary objective of this paper is to rely on experimental methods to explore and contextualize this phenomenon, ensuring greater data quality through different techniques requiring incentivized elicitation mechanisms and CE. Naturally, eliciting CE enables us to gauge the extent of the paradox. Besides, presenting the gambles individually ensures superior comprehension, considering the inherent complexity of such gambles. Aiming to address potential concerns related to comprehension issues and measurement error, we undertake the elicitation of the CE for each gamble twice, which allows us to rectify the measurement error via the ORIV technique (Gillen et al. (2019)).

Several factors might drive the Two-Ball Ellsberg Paradox, and our experimental design allows us to investigate the essential factors. One hypothesis is that subjects mistakenly think the probability of winning decreases as the ratio of red to blue balls becomes more uneven. However, our experiment shows that when subjects choose between two urns with ambiguous compositions but one more unevenly distributed one, they prefer the more unevenly distributed one. This result suggests that subjects didn't "learn" to choose the ambiguous urn even after being exposed to scenarios requiring reasoning about the ratio of red to blue balls. The preference for the risky urn is a deliberate decision to avoid ambiguity, even at a lower win probability.

Building on previous studies questioning existing models and proposing other paradoxical behaviors, we examine the relationship between the Two-Ball Ellsberg Paradox and simple behavioral mechanisms. Specifically, we explore complexity aversion by replicating Halevy's experiment, which elicits preferences between a simple lottery with a 50% win probability and a more complex compound lottery with the same probability. We also examine the relationship with classical ambiguity aversion by replicating Ellsberg's original experiment. Our results suggest a strong correlation between classical ambiguity aversion and complexity aversion.

Ambiguity and ambiguity aversion are relevant to policymakers since, in most real-world situations, agents cannot attach precise probabilities to the possible outcomes. Relying on the standard models cited above, researchers have explored the important implications of ambiguity in diverse economic fields. In environmental economics, Millner et al. (2013) demonstrate the effects of ambiguity on the social cost of carbon by integrating ambiguity within Nordhaus' famous integrated assessment model of climate economy. Lange and Treich (2008) investigate the learning effects of climate policy under ambiguity. In health economics, Treich (2010) shows under which conditions ambiguity aversion increases the value of a statistical life. In macro-finance, Ju and Miao (2012) show how ambiguity aversion can account for the equity premium puzzle. Such policy recommendations might need revisions based on updated models that accommodate our findings.

Although several experiments contain scenarios comparable to our Two-Ball gamble or draw similar conclusions to those we draw, our experiment sets itself apart by introducing a new class of Two-Ball drawings. These drawings feature ambiguity but guarantee a minimum win probability at least as large as a related non-ambiguous gamble. This design feature allows us to test whether subjects avoid ambiguity per se or avoid ambiguity due to potentially worse outcomes.

Firstly, Epstein and Halevy (2019) use a Two-Ball gamble in a supplemental treatment from a 2014 experiment. However, the authors don't elicit subjects' Certainty Equivalents for this gamble and don't observe the choice over a risky bet. Although not directly comparable, their results show that 21.6% prefer the 1-Ball ambiguous gamble over the 2-Ball ambiguous gamble among subjects with monotone and transitive choices – consistent with our findings when considering possible preference for a 50-50 risky gamble over a 1-Ball ambiguous gamble.

Fleurbaey (2017) creates a thought experiment with a risky urn (R) and an ambiguous urn (A). The decision-maker draws two balls sequentially from a combination of these urns and wins if the balls have the same color. Our paper's central Two-Ball gamble compares two draws from urn A to two from R; we do not let subjects switch urn after the first draw. While both papers explore situations where individuals may pay to avoid ambiguity, only our Two-Ball Ellsberg Paradox shows that individuals choose a dominated gamble to escape ambiguity. Moreover, Yang and Yao (2017) designed an experiment where two balls were drawn from a single urn containing red and white balls, with the payoff determined by the balls' colors. They find that up to 45% of risk-averse subjects choose urn A over urn R, violating theories that include a monotonicity axiom. These results resemble our findings, except that our central Two-Ball gamble's payoff has a mean that increases with the dispersion of the urn's contents, making urn A attractive to both risk-averse and risk-seeking individuals.

Finally, very recently, Kuzmics et al. (2020) also examined an incentivized experiment where subjects choose between a risky urn R with a known win probability of 49% and an ambiguous urn A with green and yellow balls. They find that 48.1% of subjects bet on urn R after seeing certain informational draws, which is a dominant decision strategy. However, unlike our learning treatment, they observe that "paradoxical" choices decrease in frequency after subjects are shown explanatory videos.

Our paper unfolds as follows. Section 2 outlines the experimental design and methodology employed. Section 3 shares the findings from our core gambles, spotlighting the Two-Ball Ellsberg Paradox. Section 4 delves into different hypotheses aiming to test whether participants truly understand the gambles. Section 5 explores different channels that might explain the Two-Ball Ellsberg Paradox, ranging from complexity aversion and other "paradoxical" preferences to the impact of the number or proportion of draws from ambiguous urns on participants' aversion to ambiguity, even when it can only boost their win probability. Section 6 offers concluding discussions and directions for future research to identify further channels to such a paradox.

2 Design, Data Collection and Setting

Our experiment was designed to answer two primary questions. First, to what extent do subjects prefer urn R over urn A in our Two-Ball gamble? Second, what possible explanations of this "paradoxical" preference can be falsified? Answering the first question only requires asking subjects about a few different gambles. However, since many possible explanations exist for a preference for urn R over urn A, our experiment includes many gambles designed to address the second question.

We used Prolific, an online survey platform, to run our experiment and collect our data. Due to its participant pool's quality, Prolific is increasingly used in economics to conduct surveys and incentivized experiments. Our sample comprised 880 participants, selected to be nationally representative in age and gender. Of these initial 880 participants, 708 passed the basic attention-screening questions and criteria described at the end of this section.

Due to the constraints on subjects' time and attention inherent in an online experiment, our various gambles were divided across four treatments, with each subject completing exactly one treatment. All treatments ask subjects about our central two-ball gamble (playing with the ambiguous run versus risky). All treatments elicit subjects' ambiguous attitudes via the classic two-urns Ellsberg paradox. Beyond this, each treatment contains some gambles specific to that treatment. Gambles similar to each other were grouped into blocks, and gambles within a block were presented in random order.[1]

In each gamble, the subject can either "win" (gain \$3) or "lose" (gain nothing). After viewing instructions explaining the conditions under which the current gamble will win or lose, the subject must report her certainty equivalent (CE) for that gamble from a multiple price list (MPL) containing dollar amounts between \$0 and \$3 in increments of 10 cents. Compared to eliciting choices, the MPL allows us to measure the intensity of subjects' preferences.

Laboratory and online experiments eliciting subjects' CEs for gambles are often prone to significant measurement error. To correct this, we rely on the Obviously Related Instrumental Variables (ORIV) method of Gillen et al. (2019). Compared to other methods to correct measurement errors, such as using the first elicitation as an instrument for the second, the ORIV approach generally results in lower standard errors. We, therefore, elicit subjects' CEs twice for most of our gambles.

Including all duplicate questions, each treatment contains 11 or 12 gambles in total. In each treatment, three[2] of these gambles were selected at random for incentivization: if a gamble was selected, then a random row of the MPL for that gamble was chosen, and subjects were given what they reported they preferred from that row.[3] Subjects received an average payment of \$3.50 from the incentivized questions, plus a fixed \$2 payment for completing the experiment.

Since the monetary stakes of the experiment were not very high, there is a reason for concern that subjects may answer at random to finish the experiment quickly. We employed three screening criteria to address this concern: (1) After the experiment in-

[1]The order of the blocks was also randomized; we detail the particular randomization for each treatment in Sections 4.1, 4.2, 4.3 and 5.1.

[2]Although incentivizing only one gamble would allow us to raise the monetary stakes of each question, doing so would create too large a variance in different subjects' payoffs, which was undesirable for this online experiment.

[3]For example, if Gamble X was selected for incentivization, and then the row "\$1.20" was selected at random for this gamble, the following happens. (A) If the subject reported she preferred a fixed \$1.20 payment to play Gamble X, then she received \$1.20. (B) If the subject reported she preferred playing Gamble X to receiving \$1.20, then we simulated Gamble X and gave her \$3 if it won and \$0 if it lost.

structions, but before the gambles, subjects were given a 3-question basic comprehension quiz about the instructions. Any subject who failed at least one of these questions was given a small payment and forced to leave the experiment. (2) Subjects were given a standard attention-screening question between each of the experiment's major sections. Subjects failing at least one such question were removed from our analysis. (3) If, across our two elicitations of a subject's CE for the same gamble, the subject reported two CEs that differed by more than $1 (that is, one-third the size of the $3 MPL table), that subject was removed from our analysis.[4] Out of an initial pool of 880 subjects, 172 were removed due to violating at least one of the criteria (1)-(3).

In more detail, subjects were randomly assigned to one of the following four treatments – learning, robustness, order and complexity – that we present now.

In treatment learning, subjects complete the blocks BoundedA, Ellsberg and 2Ball as well as the duplicate blocks EllsbergD and 2BallD. The order in which these blocks were presented was determined at random, independently for each subject assigned to this treatment, according to Figure 1.

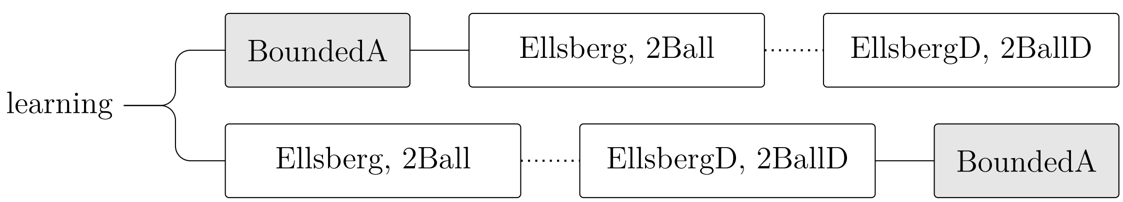

Figure 1: Structure of Treatment learning

In this figure, the initial split between line 1 (with BoundedA at the beginning) and line 2 (with BoundedA at the end) indicates that subjects were randomized uniformly between doing block BoundedA either before or after all the other blocks in the treatment. Furthermore, the fact that the boxes containing "Ellsberg, 2Ball" and "EllsbergD, 2BallD" are adjacent and shaded in the same way indicates that, within each of these two randomized groups, there is further randomization as to whether the blocks Ellsberg and 2Ball are both completed before blocks EllsbergD and 2BallD or are both completed after these two blocks. Finally, in any box containing multiple block names, those blocks were completed in a random order (e.g., block Ellsberg is either completed before or after block 2Ball). Hence, Figure 1 indicates 16 possible orders in which subjects could complete the blocks in treatment learning.

[4]Other reasonable thresholds for exclusion, such as "differed by more than $1.50," yield qualitatively similar results in our analysis as detailed in the Appendix.

In treatment robustness, subjects complete the blocks Independent, 3Ball, Ellsberg and 2Ball as well as the duplicate blocks EllsbergD and 2BallD. The order in which these blocks were presented was determined at random, independently for each subject assigned to this treatment, according to Figure 2. Its interpretation is analogous to that of Figure 1; there are 48 different orders in which the six blocks comprising Treatment robustness could be completed.

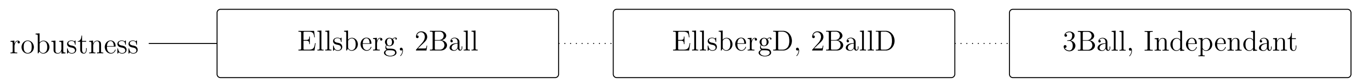

Figure 2: structure of treatment robustness

In treatment order, subjects complete the blocks 2BallMixed and Ellsberg as well as the duplicate blocks 2BallMixedD and EllsbergD. The order in which these blocks were presented was determined at random, independently for each subject assigned to this treatment, according to Figure 3. Its interpretation is analogous to that of Figure 1; there are 8 different orders in which the 4 blocks comprising Treatment order could be completed.

Figure 3: structure of treatment order

In treatment complexity, subjects complete the blocks Compound, Ellsberg and 2Ball as well as the duplicate blocks CompoundD, EllsbergD and 2BallD. The order in which these blocks were presented was determined at random, independently for each subject assigned to this treatment, according to Figure 4. Its interpretation is analogous to that of Figure 1; there are 12 different orders in which the six blocks comprising Treatment complexity could be completed. Table 6 in Appendix A.2 contains summary statistics for each elicitation of CEs for gambles C and CC.

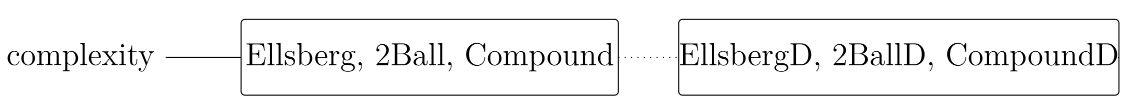

Figure 4: structure of treatment complexity

3 Participants Exhibit the Two-Ball Ellsberg Paradox

The block 2Ball contains this experiment's central gambles and is present in all four of our treatments. It contains two gambles, named RR and AA:

> RR: Draw 2 balls with replacement from urn R = [50 red, 50 blue]; win if the two balls have the same color.
>
> AA: Draw 2 balls with replacement from urn A = [Unknown red, Unknown blue]; win if the two balls have the same color.

The block Ellsberg replicates the classic Ellsberg paradox to elicit subjects' attitudes towards risk and ambiguity and is also present in all four of our treatments; it contains two gambles named R and A:

> R: Choose a color. Draw a ball from urn R = [50 red, 50 blue]; win if the drawn ball has the color you chose.
>
> A : Choose a color. Draw a ball from urn A = [Unknown red, Unknown blue]; win if the drawn ball has the color you chose.

The blocks 2BallD and EllsbergD contain duplicate gambles of those in blocks 2Ball and Ellsberg. When double-eliciting CEs, the standard practice requires the two "duplicate" gambles measuring the same CE to have slightly different wordings so that two constitute two independent measurements of that CE. To accomplish this, whenever we duplicate a block of gambles, we slightly change the specified total number of balls in a given urn without changing the proportion of balls of each color. For example, in block 2BallD, urn R contains 40 red and 40 blue balls rather than 50 red and 50 blue.

For each gamble X that is double-elicited, we use the notation X_i^j to represent the j-th elicitation of subject i's CE for gamble X, and we use the notation

$$X_i = \frac{X_i^1 + X_i^2}{2}$$

to denote the average CE of subject i for gamble X. So, for example, RR_{36}^2 represents the 2nd elicitation of subject 36's CE for gamble RR, and A_{15} denotes subject 15's average CE for gamble A. Figure 5 shows the CDFs of the empirical distributions of the CEs for RR, AA, R, and A; Table 6 in Appendix A.2 contains summary statistics for each elicitation of these CEs.

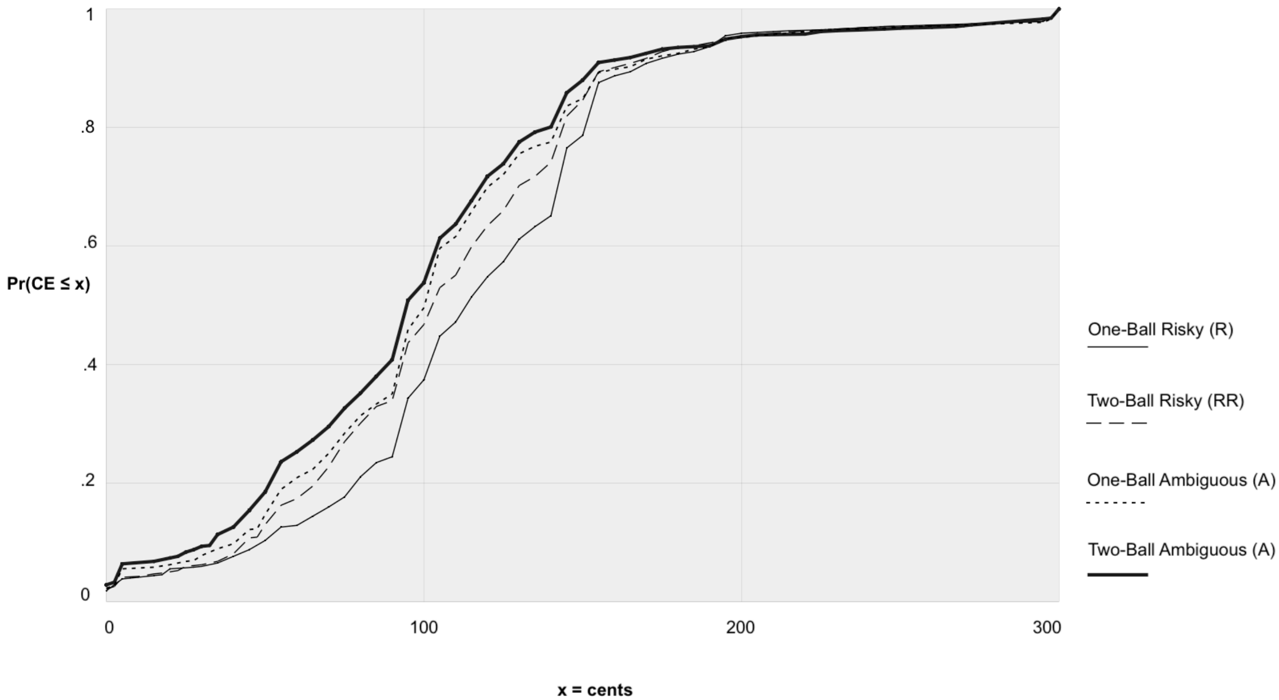

Figure 5: Cumulative Distribution of CEs for R, A, RR, AA

Other than at a few extreme CE values that were reported by a total of less than 10% of subjects, these empirical CDFs lie in the same vertical order everywhere. This suggests that on average, subjects prefer the gambles in the order $R \succ RR \succ A \succ AA$. Nearly all widely-used models of decision making under risk and ambiguity cannot explain a preference for R over AA or a preference for RR over AA, since gamble AA has a win probability of at least 50% while gambles R and RR have a win probability of exactly 50%.

Throughout this paper, we use the variable $R - AA$ to measure the extent to which individuals exhibit the "Two-Ball Ellsberg Paradox." Both $R - AA$ and $RR - AA$ are potentially useful measures of the extent to which subjects exhibit aversion to our ambiguous Two-Ball gamble AA. Indeed, $R - AA$ measures this aversion as compared to a simple 50-50 lottery, and $RR - AA$ measures this aversion as compared to a Two-Ball 50-50 lottery. Although gamble RR is mechanically more similar to AA than gamble R is, we use gamble R since it provides a more standard baseline and allows for natural comparisons with other types of aversion identified in the experimental literature. In the literature it is common to measure a subject's aversion to a newly identified phenomenon by creating some new gamble, eliciting the subject's CE X for that new gamble, and then comparing

X to that subject's CE for a simple 50-50 gamble; that is, aversion is measured with the number $R - X$. For example, when replicating Ellsberg (1961)'s experiment, classical ambiguity aversion is usually measured with $R - A$; and in Halevy (2007)'s experiment, aversion to a compound 50-50 lottery C can be measured with $R - C$.[5] Measures like $R - A$ and $R - C$ are much more naturally compared to $R - AA$ than to $RR - AA$; for a natural comparison with $RR - AA$, one would need to use strange measures like $RR - A$ and $RR - C$, which cannot even be determined from experiments where the CE for gamble RR was not measured.

Although our primary measure of the Two-Ball Ellsberg Paradox is $R - A$, our results remain qualitatively unchanged if one substitutes $RR - AA$ for $R - AA$. For example, we find that both of these variables take on a statistically significant positive value - and all the same standard models of decision making are falsified by a statistically significant positive value of $R - AA$ as would be falsified by a statistically significant positive value of $RR - AA$. With this in mind, figure 6 shows the distribution of individuals' reported CE differences $R^j - AA^j$ in each of the two elicitations j.

[5] Although C is not the notation used by Halevy (2007), we use this notation here since it is consistent with the notation introduced in Section 5.1 below.

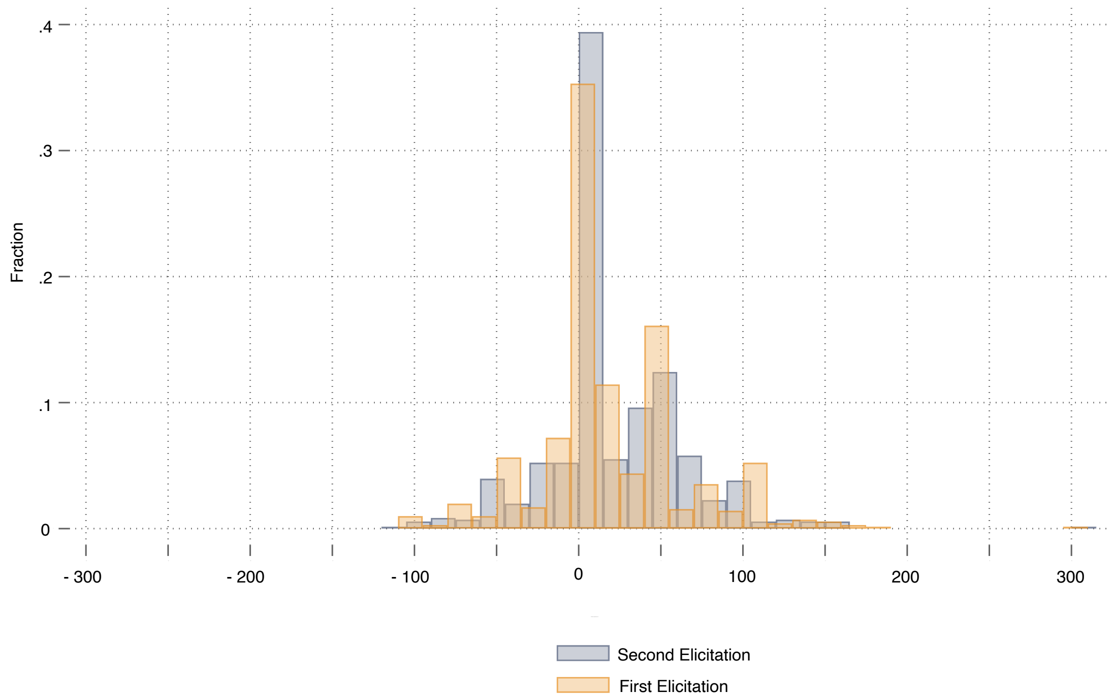

Figure 6: Histogram of $R - AA$, by elicitation

Averaging across both elicitations, a majority (54.9%) of subjects exhibited 2-Ball Ellsberg Paradox preferences by reporting a value $R - AA$ greater than zero. The average CE for gamble R is 118.13 cents, while the average CE for gamble AA is only 101.03 cents. The 17.1 cent difference between these averages is statistically significant ($t = 11.7$); individuals are willing to pay about 17% more for gamble R than they are for the higher-win-probability gamble AA.

Similarly, 44.6% of subjects prefer gamble RR over gamble AA. The average CE for RR is 109.60 cents, or 8.5% larger than AA. Its difference from AA is statistically significant ($t = 7.4$).

The next three sections explore whether these Two-Ball Ellsberg Paradox preferences are explainable solely in terms of subjects misunderstanding the gambles or otherwise maintaining false beliefs about the nature of these gambles. Sections 4.1, 4.2 and 4.3 respectively test whether subjects maintain the false beliefs Uneven is Bad, Independent Recomposition, or Dependent Recomposition mentioned in the Introduction.

4 Do Subjects Understand The Two-Ball Gamble?

There are different ways to define the notion of "comprehension" in our experimental setting. We explored three core interpretations that we report in this section. In Section 4.1, we determine whether subjects understand that the more the ambiguous urn is unevenly composed the better it is for them in terms of win probability. In Section 4.2, we check whether subjects believe that the urn contents are independently redetermined between draws or not. In the same vein, in Section 4.3, we also check whether subjects believe that the urn contents are dependently redetermined between draws or not.

4.1 Do Subjects Understand that Unevenness is Better?

Treatment learning was designed to test whether subjects behave as if they believe Uneven is Bad. The gambles unique to treatment learning are those in block BoundedA. In this block, subjects play a 2-Ball gamble: two balls are drawn from an urn A containing 100 balls, all red or blue, but whose exact contents are unknown. The subject wins \$3 if the two balls have the same color. In each gamble in block BoundedA, some further information is given about the contents of urn A, as described below.

BB^{40-60}: Urn A is known to contain between 40 and 60 red balls.

BB^{60-100}: Urn A is known to contain between 60 and 100 red balls.

BB^{95-100}: Urn A is known to contain between 95 and 100 red balls.

In treatment learning, subjects complete the blocks BoundedA, Ellsberg and 2Ball as well as the duplicate blocks EllsbergD and 2BallD. The order in which these blocks were presented was determined randomly and independently for each subject assigned to this treatment. They either faced first BoundedA and then randomly the Ellsberg Paradox and the Two-Ball Ellsberg Paradox (and their duplicate), or they started with the latter two blocks (and their duplicate) in random order and then faced BoundedA.

If subjects always believe Uneven is Bad, then we should certainly not find either of the preferences $BB^{95-100} \succ BB^{60-100}$ or $BB^{60-100} \succ BB^{40-60}$. Subjects exhibiting such preferences is evidence that we should reject the hypothesis that subjects always believe Uneven is Bad.

A subtler hypothesis to explain a preference for R over AA is that subjects believe Uneven is Bad until they are confronted with examples that demonstrate that Uneven is Good - i.e. that a more uneven urn yields a higher win probability in a 2-Ball gamble. For example, subjects may believe Uneven is Bad when asked "out of the blue" about gamble AA, but may come to believe Uneven is Good only after considering e.g. gamble BB^{95-100} and realizing that an urn containing at least 95% red balls is very likely to lead to a win. We call this the Learning Hypothesis, as it entails that subjects are "nudged" into believing that Uneven is Good when exposed to certain suggestive 2-Ball gambles.

A preference $BB^{95-100} \succ BB^{60-100} \succ BB^{40-60}$ is consistent with the Learning Hypothesis since subjects may be "nudged" into the belief Uneven is Good as early as the beginning of block BoundedA. However, if the Learning Hypothesis is true, then subjects in treatment learning should report a smaller average value of $R - AA$ than those in other treatments - since only those subjects in treatment learning had any exposure to block BoundedA.

Figure 7 shows the CDFs of the empirical distributions of the CEs from treatment learning for gambles BB^{40-60}, BB^{60-100} and BB^{95-100}. It also shows the combined CDF (from all 4 treatments) of CEs for gamble AA. Table 1 gives summary statistics of BB^{40-60}, BB^{60-100} and BB^{95-100} as well as for AA using only those subjects in treatment learning.

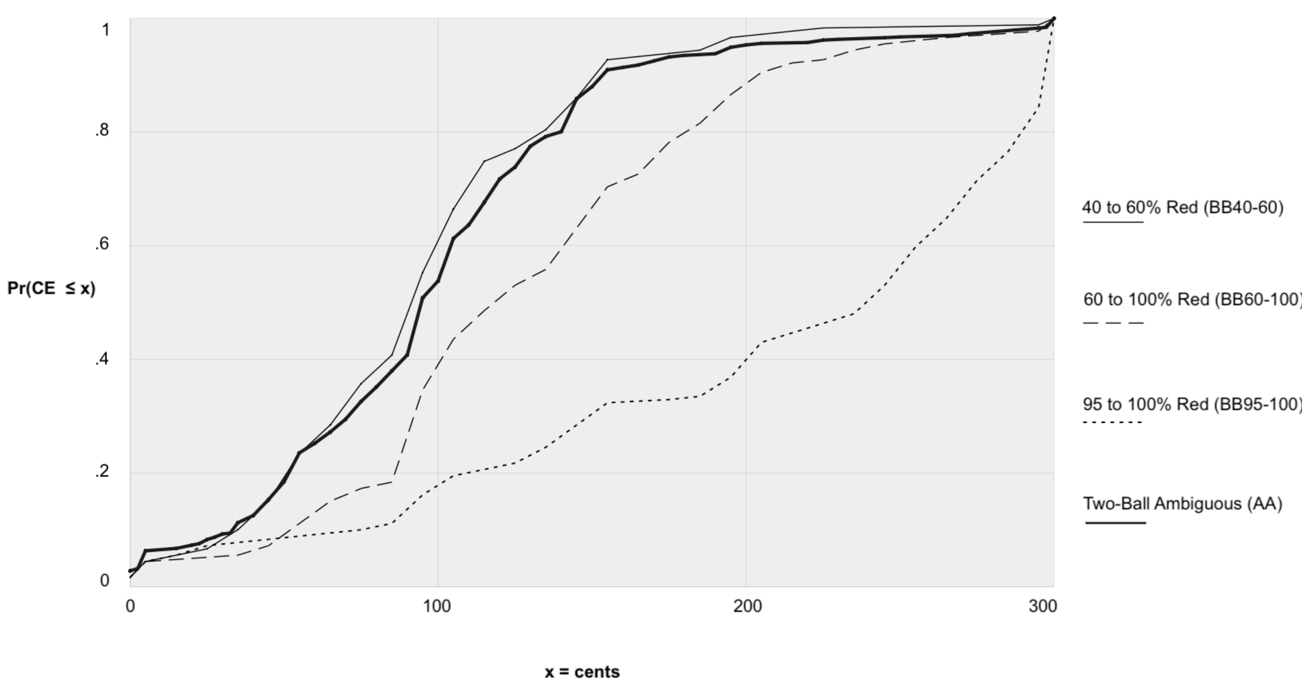

Figure 7: Cumulative Distribution of CEs for BB^{40-60}, BB^{60-100}, BB^{95-100}, AA.

	BB^{40-60}	BB^{60-100}	BB^{95-100}	AA
Mean	98.603	132.235	207.654	97.179
SD	(52.113)	(63.887)	(90.784)	(55.634)
N	179	179	179	179

Table 1: CEs for Subjects in Treatment learning

Table 1 suggests that subjects do not always believe Uneven is Bad. Although there is no statistically significant difference between the average CE for AA and that of BB^{40-60}, subjects prefer BB^{60-100} to BB^{40-60} by an average of 33.6 cents ($t > 9$). Similarly, they prefer BB^{95-100} to BB^{60-100} by an average of 75.4 cents ($t > 13$). Of the 179 subjects in treatment learning, 147 reported the "correct" ranking $BB^{95-100} \succsim BB^{60-100} \succsim BB^{40-60}$. Only 22 of the 179 subjects reported a larger CE for BB^{40-60} than BB^{60-100}; and even among those 22 subjects, the average CE for BB^{95-100} was massively larger than the average CE for BB^{60-100} (mean of difference = 68.64, $t = 3.36$). These data suggest we must reject the hypothesis "Subjects always believe Uneven is Bad" as an explanation for subjects' behavior.

Now consider the Learning Hypothesis. If this latter is true, then we should find that the CE difference $R - AA$ is significantly smaller (or, more negative) among subjects who completed block BoundedA before completing blocks 2Ball and 2BallD than it is among subjects who did not complete BoundedA before 2Ball and 2BallD. Completing block BoundedA should "nudge" subjects into being less susceptible to the 2-ball Ellsberg paradox.

Half of the 179 subjects randomly assigned to treatment learning completed BoundedA before the blocks 2Ball and 2BallD, whereas none of the subjects randomly assigned to other treatments did so. So if the Learning Hypothesis is true, we should find a statistically significant (negative) difference between the $R - AA$ values in the treatment learning versus those in the other treatments.

If we let I^{TN} be the indicator variable for assignment to Treatment learning, then in a regression of $Z := R - AA$ on I^{TN}, the slope coefficient represents the causal effect of being in treatment learning on the preference for R over AA. A statistically significant negative slope coefficient would be evidence that the Learning Hypothesis is true.

	Z^1	Z^2	Z^{avg}
I^{TN}	2.615	-0.418	1.098
	(3.659)	(3.917)	(3.366)
Const.	16.994	16.786	16.890
	(1.840)	(1.969)	(1.692)
N	708	708	708

Table 2: learning Effects

Table 2 shows the results of such a regression, first using individual elicitations and then the averages across elicitations. As shown, the slope coefficient is not statistically significant, and it is positive in the case using averages. Thus, we fail to reject the null hypothesis ($p = .63$) and hence have no evidence of the Learning Hypothesis. The results of Treatment learning therefore provide strong evidence that a belief in Uneven is Bad - even a belief in Uneven is Bad that could be eliminated by "learning" - does not drive the Two-Ball Ellsberg Paradox.

4.2 Do Subjects Believe Urn Contents are Independently Redetermined Between Draws?

The gambles unique to treatment robustness are those in blocks Independent and 3Ball. Block Independent from treatment robustness was designed to test whether subjects behave as if they believe Independent Recomposition is true. Meanwhile, block 3Ball contains gambles designed to explore how the "amount" of ambiguity present in a gamble affects subjects' preferences; it is discussed in Section 5.2 below.

In block Independent, there is only one gamble, IA, where subjects draw a ball from each of two ambiguous urns (containing only red and blue balls) whose contents were determined independently; they win \$3 if the two balls have the same color.

In block 3Ball, subjects draw 3 balls in total, with replacement, from some combination of a single ambiguous urn A and a single risky urn R, in a certain order. They win \$3 if all three balls have the same color. We summarize the gambles below:

RRR: 1st ball from urn R; 2nd ball from urn R; 3rd ball from urn R.

AAA: 1st ball from urn A; 2nd ball from urn A; 3rd ball from urn A.

RAA: 1st ball from urn R; 2nd ball from urn A; 3rd ball from urn A.

In treatment robustness, subjects complete the blocks Independent, 3Ball, Ellsberg and 2Ball as well as the duplicate blocks EllsbergD and 2BallD. The order in which these blocks were presented was determined randomly and independently for each subject assigned to this treatment.[6]

If subjects believe in Independent Recomposition, then gamble AA should (according to them) be identical to gamble IA. We should therefore find no difference between their average CEs for gambles AA and IA. In reality, for any procedure generating the contents of ambiguous urns, gamble AA must have at least as large of a win probability as gamble IA, and AA must have a larger win probability than IA if the procedure is nondegenerate (i.e., assigns a nonzero probability to at least two different possible urn compositions).[7] Finding a preference $IA \succ AA$ would therefore be evidence in favor of the hypothesis that subjects believe in Dependent Recomposition; specifically, it is consistent with them believing that ambiguous urns' contents are recomposed adversarially between draws (i.e., the contents of ambiguous urns are chosen based on the results of draws so far and in such a way as to lower subjects' chances of winning). Conversely, finding a preference $AA \succ IA$ would be evidence consistent with subjects correctly believing that the win probability of AA is larger than that of IA and/or believing in beneficial Dependent Recomposition.

Figure 9 shows the CDF of the empirical distribution of CEs from gamble IA - the only gamble in block Independent. For comparison, it also shows the combined CDF (from all 4 treatments) of CEs for gamble AA.

The mean CE for gamble IA was 107.839, and the standard deviation of these CEs was 68.733. On average, the 192 subjects in treatment robustness slightly preferred AA to IA, but the difference is not statistically significant (mean = 1.73, $t = .60$).

We therefore have insufficient evidence to reject the hypothesis that subjects believe in Independent Recomposition. A future experiment that replicates block Independent with a larger sample size or larger payments may be able to reject this hypothesis; see

[6]This section explores the results from block Independent. We discuss block 3Ball in Section 5.2 since this block was designed to address very different hypotheses from those currently being discussed.Block 3Ball was included in treatment robustness due to the time constraints of our online experiment.

[7]If the procedure is symmetrical, i.e. for any $x \in [0, .5]$ it is just as likely to have exactly a $.5 + x$ proportion of red balls as it is to have a $.5 - x$ proportion of red balls, then clearly gamble IA has a win probability of exactly 50% while gamble AA has a win probability of 50% only if the procedure is degenerate (and otherwise has a larger win probability). If the procedure is not symmetrical then gamble IA will have a win probability larger than 50%, but that of AA will be larger still. For example, if the procedure is "with probability .5 we make the ambiguous urn contain 50% red balls, and with probability .5 we make it contain 100% red balls," then gamble IA has win probability .625. In contrast, gamble AA has a win probability .75.

also Section 6 for discussion of a variation on block BoundedA that may be able to reject this hypothesis in a future experiment.

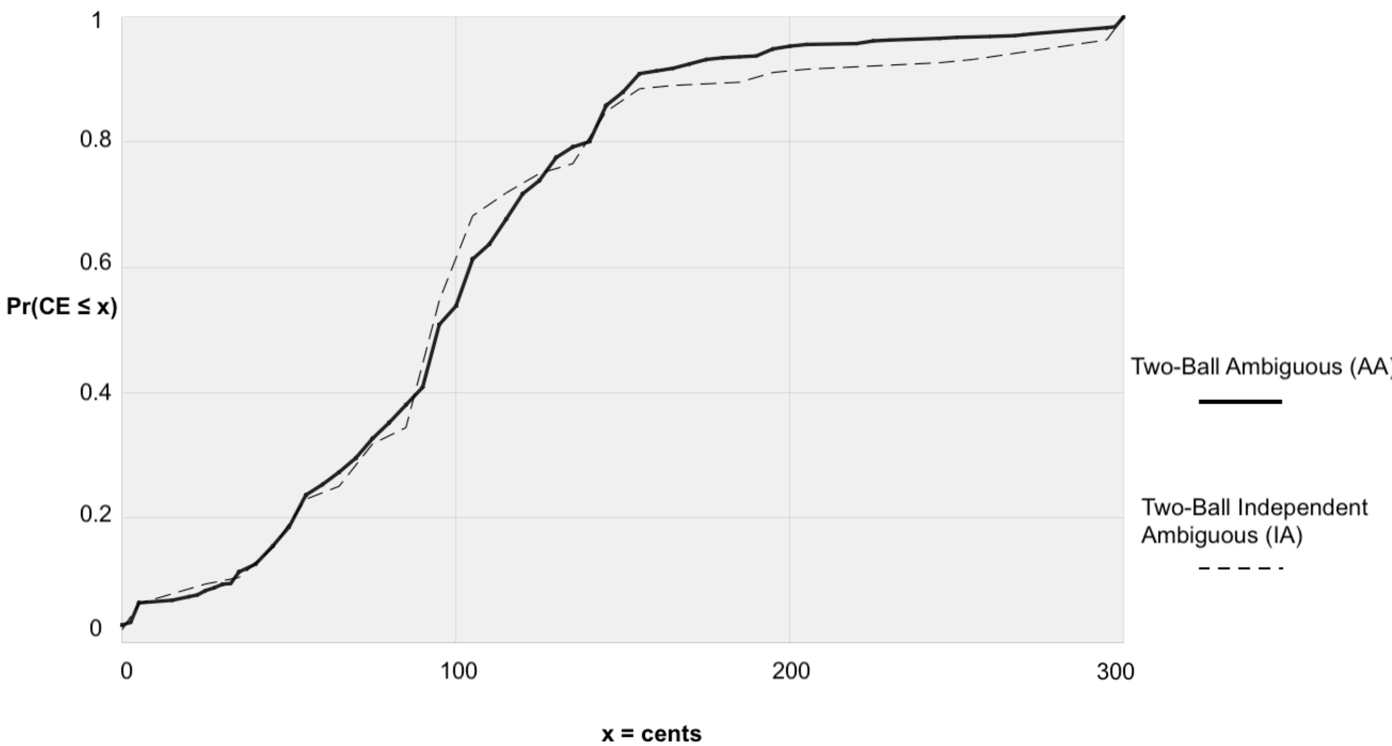

Figure 8: Cumulative Distribution of CEs for AA, IA

Since the average CE for gamble AA was slightly larger than that of gamble IA, treatment robustness provides no evidence that subjects believe in adversarial Dependent Recomposition. Treatment order was designed to more generally test whether subjects believe in Dependent Recomposition in any form, either adversarial or beneficial; as we will see in Section 4.3, our findings there similarly provide no evidence of belief in Dependent Recomposition.

This means that subjects' preference for gamble R over gamble AA is unlikely to be due to a false belief that gamble AA has lower win probability because urn A's contents are adversarially redetermined between draws.

4.3 Do Subjects Believe Urn Contents are Dependently Redetermined Between Draws?

Treatment order was designed to test whether subjects behave as if they believe Dependent Recomposition is true. The gambles unique to treatment order are found in block 2BallMixed - an expanded version of block 2Ball that contains gambles not only RR and AA as before but also gambles AR and RA. In each gamble, subjects draw two balls - either from the same urn and with replacement or from distinct urns - in a certain order, and they win \$3 if the two balls have the same color. We summarize these gambles below:

RR: 1st ball from urn R; 2nd ball from urn R.

AA: 1st ball from urn A; 2nd ball from urn A.

AR: 1st ball from urn A; 2nd ball from urn R.

RA: 1st ball from urn R; 2nd ball from urn A.

Block 2BallMixedD contains duplicate questions of those in block 2BallMixed. In treatment order, subjects complete the blocks 2BallMixed and Ellsberg as well as the duplicate blocks 2BallMixedD and EllsbergD. The order in which these blocks were presented was determined randomly and independently for each subject assigned to this treatment.

If subjects believe in Dependent Recomposition, then they should report different average CEs for gamble RA than they report for gamble AR. Indeed, whether subjects believe the recomposition of urn A's contents is done adversarially or beneficially, gamble AR must have a win probability of exactly 50% since, whatever ball was drawn from the first urn, there is a 50% chance of drawing a ball of that color from urn R in the second draw. Meanwhile, if adversarial (beneficial) Dependent Recomposition is true, then gamble RA has a win probability that is smaller (greater) than 50%.

Conversely, if Dependent Recomposition does not hold, then gambles AR and RA both have a win probability of 50%. Finding that $AR \sim RA$ is therefore evidence that subjects do not believe in Dependent Recomposition.

In reality, gambles RA and AR both have a win probability of exactly 50%, while gamble AA must have a win probability of at least 50%. Subjects exhibiting preferences $AA \succ AR \sim RA \sim RR$ would be consistent with them fully understanding these win probabilities and basing their preferences on nothing but these win probabilities.

A preference $RR \succ AA$ would suggest that, if subjects understand the win probabilities of these gambles, then they must harbor a distaste for either the mere presence of ambiguity in a gamble or for the amount of ambiguity present in a gamble (as measured by the number or proportion of draws that come from ambiguous urns). Our results from Section 4.1 strongly suggest that subjects understand that gamble AA has a win probability that is larger than 50% (and hence larger than the win probability of RR), but they do not directly imply that subjects understand the win probabilities of gambles AR and RA to be exactly 50%.

Assuming subjects understand the win probabilities of these gambles, preferences $RR \succ AA \succ AR \sim RA$ are consistent with subjects harboring a distaste for either the mere presence or the amount of ambiguity in a gamble, but preferences $RR \succ AA \sim AR \sim RA$ are consistent only with subjects harboring additional distaste based on the amount of ambiguity in a gamble. Indeed, both gambles AA and RA have ambiguity present, but gamble AA has a larger win probability; hence an indifference between them implies that an additional distaste for the second ambiguous draw must be offsetting the increased win probability of gamble AA.

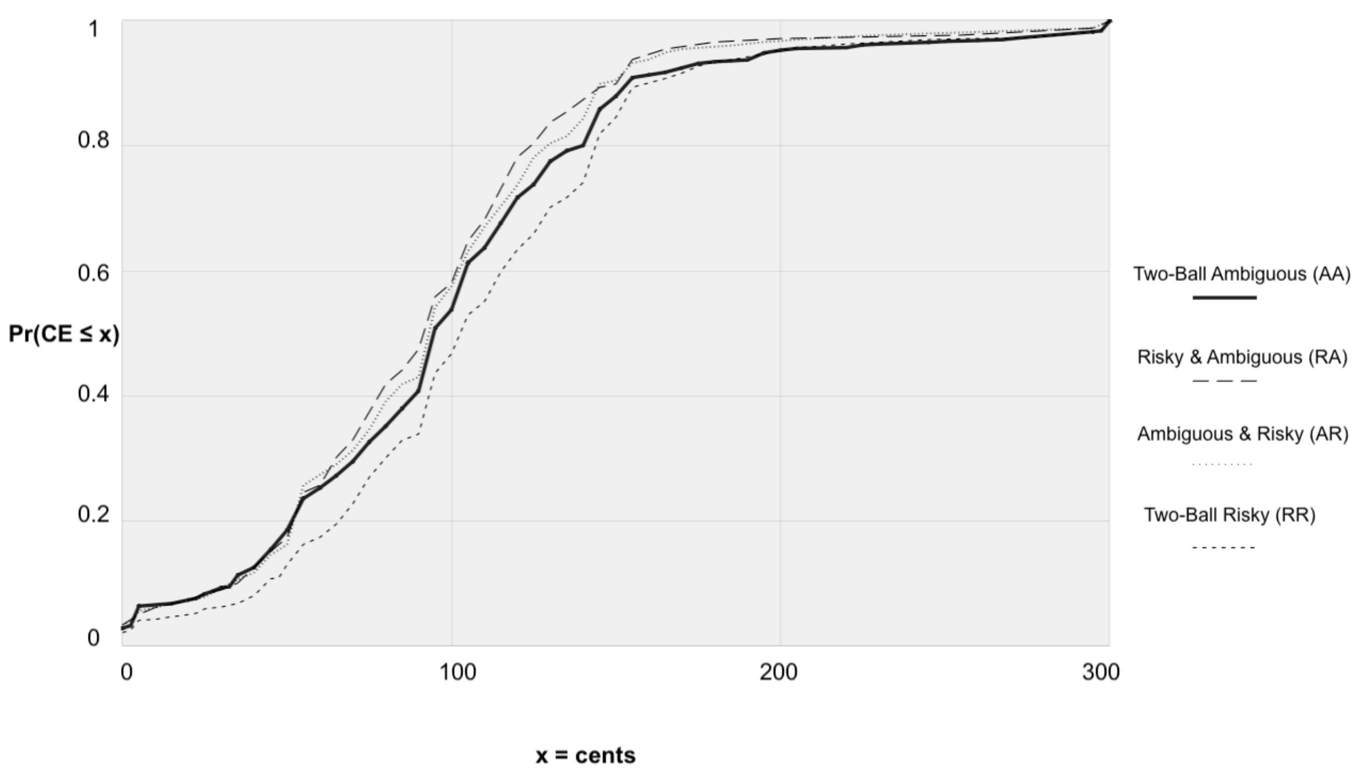

Figure 9: Cumulative Distribution of CEs for AA, RA, AR, RR

Figure 9 shows the CDFs of the empirical distributions of the CEs from treatment order for gambles AR and RA. For comparison, it also shows the combined CDFs (from all 4 treatments) of the CEs for gambles AA and RR.

Table 6 in Appendix A.2 contains summary statistics for each elicitation of CEs for gambles RR, AA, AR, and RA. Table 7 in Appendix A.3 contains summary statistics for each elicitation of the differences between the CEs RR, AA, AR, and RA.

The average CEs for the four gambles in block 2BallMixed are ranked in the order

$$RR > AA > AR > RA,$$

but the only statistically significant differences between these variables are those between RR and each of the other three. Hence, we writing an indifference wherever we cannot rule one out, subjects' preferences are of the form

$$RR \succ AA \sim AR \sim RA. \tag{i}$$

Since $AR \sim RA$, we fail to reject the null hypothesis and hence we have no evidence that subjects believe in Dependent Recomposition.

Since $RR \succ AA$, we cannot conclude that subjects base their preferences entirely on the (true) win probabilities and nothing else. This finding is consistent with the preference $R \succ AA$ observed across all treatments.

The indifference $AA \sim AR$ may be due to a false belief in Independent Recomposition. As discussed in Section 4.2, we lack sufficient evidence to rule out this hypothesis. However, our results from Section 4.1 suggest that subjects largely understand the win probabilities of 2-Ball gambles, making this Independent Recomposition hypothesis less likely.

Assuming subjects understand the win probabilities of 2-Ball gambles, the indifference $AA \sim AR$ suggests that subjects harbor an additional distaste for each additional draw that comes from an ambiguous urn, rather than a constant level of distaste once ambiguity is involved at all. We designed Block 3Ball from Treatment robustness, discussed in Section 5.2 below, to further assist us in determining whether additional draws from ambiguous urns (even when they only improve win probabilities) make gambles less preferable.

Overall, what can explain the preference for gamble R over gamble AA? The results of treatment order provide no evidence that subjects believe in Dependent Recomposition, and even if subjects fully believe in Independent Recomposition, such a belief is not itself sufficient to produce a preference for gamble R over gamble AA. We conclude that subjects harbor a distaste for the mere presence of ambiguity, or more likely for the amount of ambiguity present, in a gamble. Section 5.1 explores whether such a distaste for ambiguity is equivalent to a distaste for complexity.

5 How Can We Explain The Two-Ball Ellsberg Paradox?

Having established that the Two-Ball Ellsberg Paradox is not entirely due to misunderstanding, this result leaves open how to explain it. We believe that there may be several channels that might explain it and our experimental design allowed us to several of them. In Section 5.1, we investigate the extent to which ambiguity aversion can be considered a form of complexity aversion. In Section 5.2, we explored whether we could define "an amount" of ambiguity by designing a "Three-Ball Ellsberg" gamble and whether such an 'amount' matters to explain the Two-Ball Ellsberg Paradox.

5.1 Ambiguity Aversion as a Form of Complexity

One might argue that the preference for gamble R over gamble AA is not due to an aversion to the ambiguity present in gamble AA but instead to the complexity present in gamble AA. "Complexity" is a concept difficult to define precisely, and it is not the aim of this paper to do so. However, experiments like Halevy (2007)'s have established the potential relevance of specific types of complexity, such as the compoundness of lotteries. With this in mind, we test whether the preferences for gamble R over gamble AA is indistinguishable from the preference for a simple 50-50 gamble like R over a compound 50-50 gamble, call it C as described in Table ??. We designed Treatment complexity to test whether these specific types of complexity may be the primary factors generating the Two-Ball Ellsberg paradox.

The gambles unique to treatment complexity are those in block Compound. In this block, subjects play two gambles involving an urn C containing 100 balls, all red or blue. Subjects are informed that before each gamble begins, the contents of urn C are determined uniformly at random (i.e., each of its 101 possible balls compositions is equally likely to be realized). We summarize these gambles below.

C: Choose a color. Draw one ball from urn C; win if it's the color you chose.

CC: Draw two balls with replacement from urn C; win if they're the same color.

In other words, block Compound consists of two gambles: a compound lottery C and a "Two-Ball Compound" gamble CC. Gamble CC is the same as the ambiguous gamble

AA, except its urn's contents are determined by a known lottery rather than an unknown, ambiguous procedure.

Block CompoundD contains duplicate questions of those in block Compound. In treatment complexity, subjects complete the blocks Compound, Ellsberg and 2Ball as well as the duplicate blocks CompoundD, EllsbergD and 2BallD. The order in which these blocks were presented was determined randomly and independently for each subject assigned to this treatment.[8]

Figure 10 shows the CDFs of the empirical distributions of the CEs from treatment complexity for gambles C and CC. For comparison, it also shows the combined CDFs (from all 4 treatments) of the CEs for gambles AA, R, and RR.

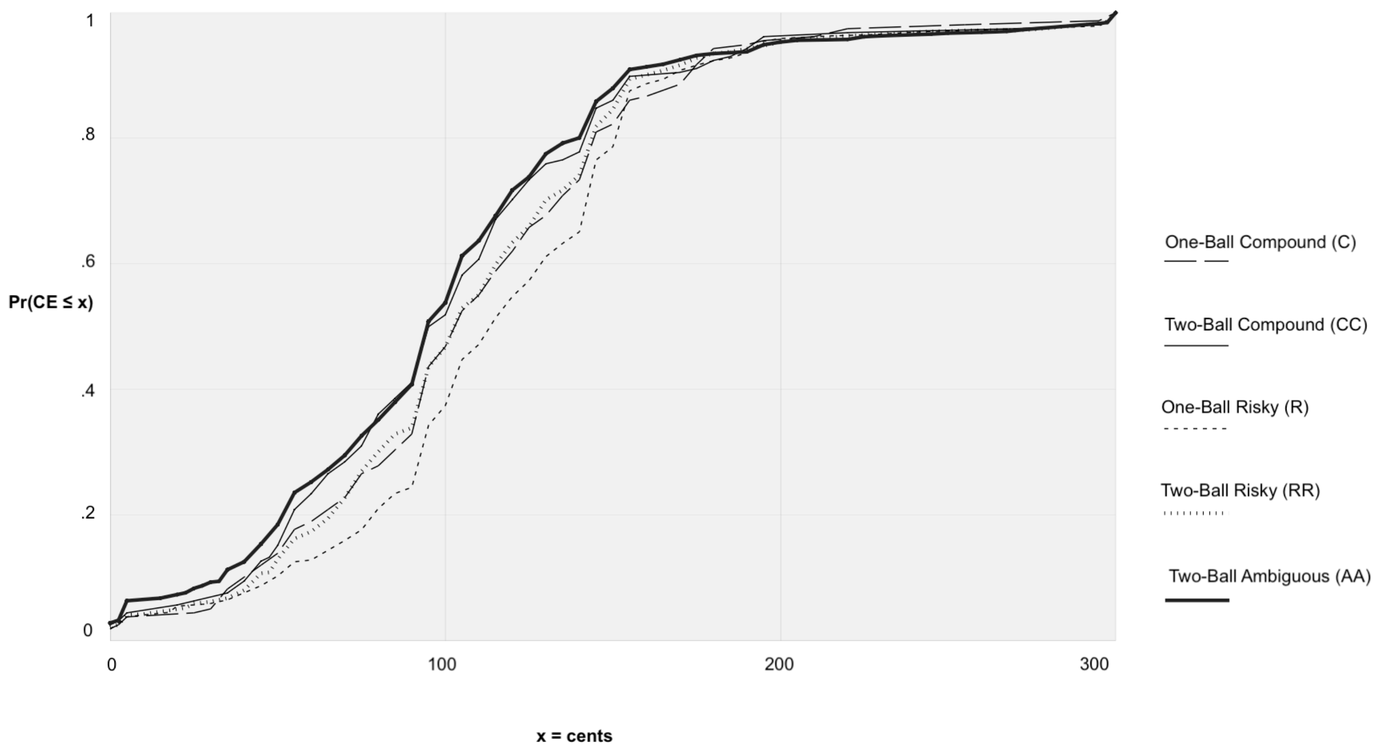

Figure 10: Cumulative Distribution of CEs for C, CC, R, RR, AA

The variable $R - A$ measures subjects' ambiguity aversion in the classic Ellsberg paradox, while $R - RR$ measures their preference for a one-ball 50-50 gamble to a Two-Ball

[8]Table 6 in Appendix A.2 contains summary statistics for each elicitation of CEs for gambles C and CC.

50-50 gamble. $R - C$ measures subjects' preference for a one-ball 50-50 gamble over a Compound 50-50 gamble, and $R - CC$ measures their preference for a one-ball 50-50 gamble over a Two-Ball Compound 50-50 gamble. Table 7 in Appendix A.3 contains summary statistics for each elicitation of these CE differences.

Table 3 computes the ORIV-adjusted correlations[9] between our central variable $R - AA$ and these other variables.

Dependent Variable: $R - AA$				
Indep. Variable:	$R - A$	$R - RR$	$R - C$	$R - CC$
ORIV ρ	0.892	0.952	0.954	0.917
	(0.017)	(0.012)	(0.024)	(0.032)
N	708	708	158	158

Table 3: relationships between ce differences

As the table shows, the preference for R over AA is extremely tightly correlated with each of the preferences mentioned in the previous paragraph. Thus, from the analyst's point of view, a subject exhibiting one of these "paradoxical" preferences to a certain degree of strength (as measured by standard deviations above the population mean) makes it exceedingly likely that she will exhibit these other "paradoxical" preferences to a similar degree of strength. In particular, this finding replicates Halevy (2007)'s and Gillen et al. (2019)'s conclusions that ambiguity aversion in the classic Ellsberg paradox is tightly linked to failure to reduce compound lotteries.

Besides correlations, it is worthwhile to examine the differences between the variables in the table above. $R - AA$ is larger than all of $R - A$, $R - RR$, and $R - C$ ($t > 4$ in all cases) and is larger than $R - CC$ by a statistically insignificant amount ($t = 1.05$). This suggests that, according to most subjects, gamble AA is likely the "worst" of gambles AA, A, RR, C, and CC - perhaps because gamble AA combines ambiguity and Two-Ball complexity. The only possible competitor for being the "worst" is gamble CC, which is identical to gamble AA except that its urn's contents are determined randomly rather than in an ambiguous manner.

[9] ORIV corrects for measurement error. If one does not do so, computed correlations are biased towards 0. Hence, these ORIV-corrected correlations may appear larger than correlations typically computed in other studies.

5.2 The "Amount" of Ambiguity Matters

Block 3Ball, a part of treatment robustness, was designed to test whether subjects exhibit a constant distaste for the mere presence of ambiguity in a gamble versus whether subjects exhibit an additional distaste for larger amounts of ambiguity in a gamble (as measured by the number or proportion of draws that come from ambiguous urns). The gambles in block 3Ball were summarized above in Table ??.

If (as suggested by our results from Section 4.1), subjects understand the win probabilities of the gambles, then a preference $RAA \succ AAA$ or $RAA \sim AAA$ indicates that the additional amount of ambiguity present in gamble AAA makes it less preferable and offsets its increased win probability, such that gamble RAA becomes at least as desirable as AAA.

Lastly, block 3Ball allows us to compare $RRR - RAA$ to $RR - AA$ to see whether it is the total number of draws that are from ambiguous urns or instead the proportion of draws that are from ambiguous urns is key to subjects' distaste for ambiguous draws. Indeed, both gambles RAA and AA feature exactly two draws from ambiguous urns, but gamble AA has all its draws from ambiguous urns while gamble RAA merely has two-thirds of its draws from ambiguous urns. 3-Ball gambles have lower win probabilities than 2-Ball gambles; for example, gamble RRR has half the win probability of gamble RR. Nonetheless, subjects' CEs for gamble RRR need not be precisely half the size of their CEs for gamble RR. Thus, to compare subject i's CE from a 2-Ball gambles to her CE from an analogous 3-Ball gamble, we first must multiply her 2-Ball CE by the factor RRR_i/RR_i. With this in mind, if we let

$$X_i = \frac{RRR_i}{RR_i} \cdot (RR_i - AA_i) - (RRR_i - RAA_i)$$

then observing a statistically significant positive average value of X indicates that a larger proportion of ambiguous draws is distasteful (holding constant the number of ambiguous draws).

Figure 11 shows the CDFs of the empirical distributions of the CEs for gambles RRR, AAA, and RAA from treatment robustness. Table 4 presents summary statistics of these CEs.

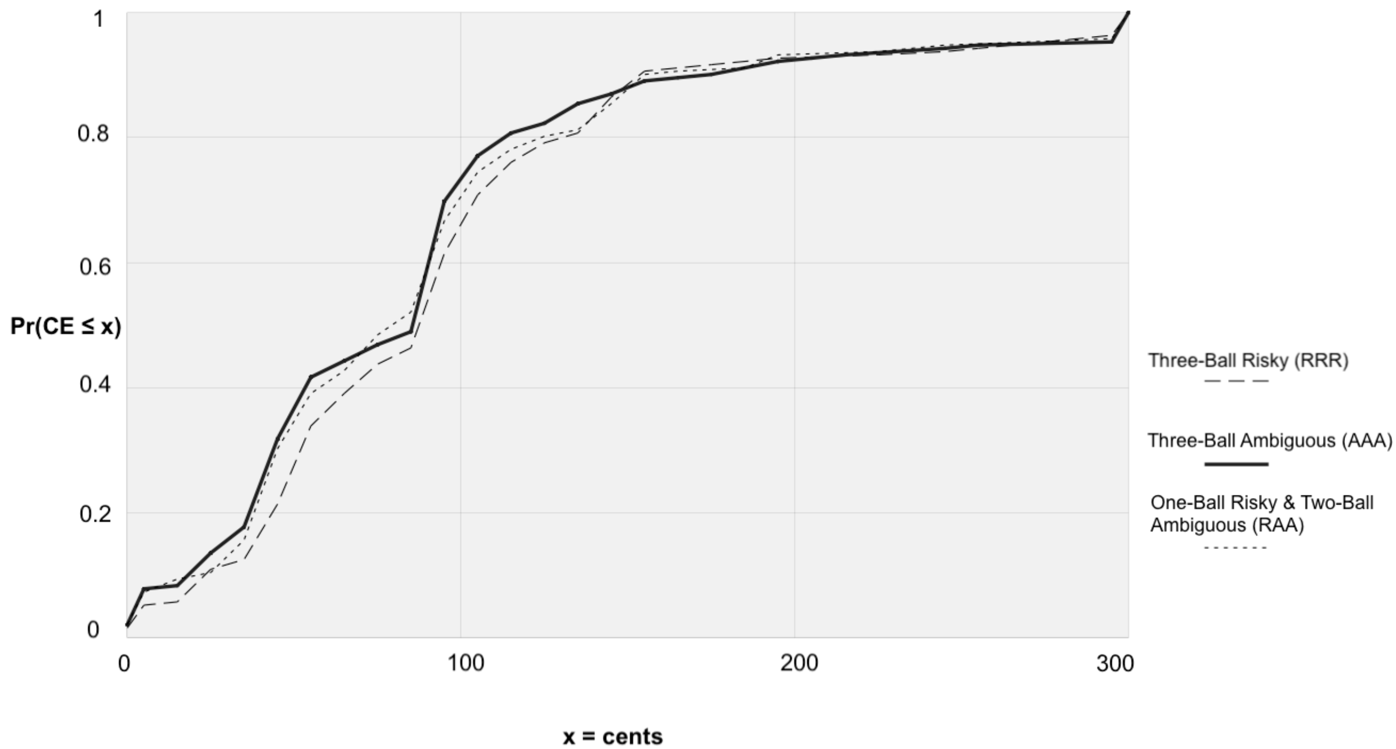

Figure 11: Cumulative Distribution of CEs for RRR, AAA, RAA

	RRR	AAA	RAA
Mean	97.708	91.120	92.552
SD	(67.310)	(69.264)	(68.172)
N	192	192	192

Table 4: CEs for 3-Ball gambles

Several striking features are apparent in these data. First, these reported CEs are too large for a classical risk-averse agent who correctly calculates the probabilities of winning.[10] Notice that gamble RRR has a win probability of exactly $\frac{1}{4}$, but subjects report an average CE of 97.7 cents for it - a value significantly larger than the risk-neutral CE of 75 cents ($t = 4.67$). Similarly, subjects on average value gamble RAA

[10]Our results from the simple 50-50 gamble in block Ellsberg suggest that subjects are on average slightly risk averse.

at significantly more than half as much as gamble AA (difference of means = 37.77, $t = 11.03$). Thus, subjects seemingly overweight the win probabilities of 3-Ball gambles.

Next, despite the general overweighting of win probabilities, comparisons between CEs for these 3-Ball gambles remain qualitatively similar to the comparisons between the CEs for Two-Ball gambles. Similarly to how subjects on average preferred RR to AA, we find that subjects on average prefer RRR to AAA (mean = 6.59, $t = 2.16$), even though AAA must have at least as large of a win probability as RRR.

On average, subjects reported a slight preference for gamble RAA over gamble AAA; however, this difference was not statistically significant (mean = 1.43, $t = .56$). As indicated above, this is consistent with a distaste for additional amounts of ambiguity in a gamble, as measured by either the number or proportion of draws that come from ambiguous urns.

The average value of the variable X defined above was negative and not statistically significant ($t = -.76$).[11] This indicates that, in terms of subjects' distaste for the presence of ambiguity, the proportion of draws that come from ambiguous urns is less relevant than the total number of draws that come from ambiguous urns.

6 Concluding Remarks

Two-Ball gambles are a rich class of decision problems. Because they can involve ambiguity but guarantee a minimum win probability that is at least as large as that of some other gamble, they allow us to test whether subjects avoid ambiguity per se as opposed to avoiding ambiguity because it may yield a worse outcome.

The most striking case of preferring a gamble with lower win probability is that subjects preferred the 50-50 gamble R to the Two-Ball ambiguous gamble AA. This preference is closely correlated with the traditional Ellsberg preference for R over a 1-Ball ambiguous gamble A, and also with the preference for R over the compound 50-50 gamble C, as well as the preference for R over the Two-Ball 50-50 gamble RR. These close relationships suggest that it may be difficult to separate an aversion to ambiguity per se from an aversion to complexity.

It is implausible that subjects prefer R to AA simply due to a poor understanding of Two-Ball gambles. In the block BoundedA, subjects correctly and strongly identified that more unevenly distributed urns are more likely to win. Moreover, the lack of a "learning"

[11] Constructing the variables X_i required us to drop those 4 subjects who, in both elicitations, reported a CE for gamble RR equal to 0. Leaving these subjects in the data set would lead to division by 0. Hence, this t-test was run with $n = 188$ rather than $n = 192$.

effect from being in the treatment containing block BoundedA suggests that even without any additional examples or explanations, subjects understand 2-Ball gambles enough to make reasonably accurate comparisons of their win probabilities.

Subjects exhibit a preference to avoid the mere presence of ambiguity in a gamble. Using the number of balls drawn from ambiguous urns as a coarse measure of the "amount" of ambiguity in a gamble, subjects seem to exhibit a stronger distaste for gambles with larger amounts of ambiguity. Further models and experiments are needed to determine the manner in which people react to situations involving various types of ambiguity.

In exploring what can explain the Two-Ball Ellsberg Paradox, treatments learning and order show that there is no evidence of subjects holding false beliefs Uneven is Bad or Dependent Recomposition. However, in block Independent from treatment robustness, we failed to find sufficient evidence to reject the hypothesis that subjects wrongly believe in Independent Recomposition. Later in this section, we discuss how a future experiment might more easily falsify the hypothesis that subjects believe in Independent Recomposition.

Even if subjects maintain some belief in Independent Recomposition, such a belief alone is not sufficient to generate a preference for gamble R over gamble AA. Indeed, even under Independent Recomposition, gamble AA must still have a win probability of at least 50%. This preference suggests that individuals harbor a distaste for the mere presence of ambiguity in a gamble.

In exploring whether the Two-Ball Ellsberg relates more to an aversion to complexity or to ambiguity and whether such a distinction, in treatment complexity, we found that the "Two-Ball Ellsberg Paradox" preference for gamble R over gamble AA was tightly correlated with other "paradoxical" preferences such as aversion to the complexity present in compound lotteries. Although the magnitude of $R - AA$ was larger than the magnitudes of nearly all of these other preferences, one might nonetheless argue that the preference for R over AA is due to a distaste for complexity rather than ambiguity.

Even in this case, we have identified the mere presence of ambiguity as a driver of change in people's behavior, perhaps through the complexity it introduces or perhaps through other means. Whether explained as an instance of complexity or not, people harboring a distaste for the mere presence of ambiguity has potentially widespread implications for economics. Subjects may prefer to gamble R to A in the classic Ellsberg paradox primarily because they dislike the mere presence of ambiguity and not, for instance, entirely because they hold concern for worst-case scenarios, as Gilboa and Schmeidler (1989) and many other models would suggest. Models ignoring a distaste for ambiguity per se would incorrectly predict individuals' behavior in a variety of situations. Hence, new models may be required.

Besides, unlike in the original Ellsberg paradox, a subject cannot eliminate the ambiguity present in gamble AA by introducing randomization in her choice of color (as in Raiffa (1961)). Indeed, gamble AA does not ask subjects to choose a color. Even if we presented subjects with a modified version of gamble AA wherein they choose either red or blue and win if and only if both balls drawn were of the chosen color (and compared this to a similarly modified version of gamble RR), it is still the case that randomizing one's color choice does not eliminate the ambiguity in the payoff of gamble AA. If p is the (ambiguous) proportion of red balls in urn A, then this modified version of gamble AA has win probability p^2 when you bet on red and win probability $(1-p)^2$ when you bet on blue.

Randomizing your choice of color 50-50 would thus mean that the gamble's win probability is $.5p^2 + .5(1-p)^2 \geq .25$. In contrast, the modified version of gamble RR has a $.25$ probability of winning, regardless of the color on which you bet (or whether you randomized your choice of color). It is still the case that gamble AA has an ambiguous win probability and that it is at least as large as (and in all but one case, strictly larger than) that of RR.

Finally, we might imagine a further experiment to reject the independent recomposition hypothesis. Recall the Independent Recomposition hypothesis mentioned in Section 4.2: Do subjects imagine that our "two draws with replacement from the same ambiguous urn" are actually "two draws from two ambiguous urns whose contents were determined independently"? Our experiment can't rule out a belief in Independent Recomposition as a partial driver of the 2-Ball Ellsberg paradox, but here we suggest how a further experiment might do so.

A variation on block BoundedA may be sufficient to show that subjects do not believe in Independent Recomposition. Consider a version of gamble BB^{95-100} wherein instead of the gamble specifying that the urn contains between 95 and 100 red balls, it merely specifies that at least 95 of the 100 balls in the urn are of the same color. Suppose subjects imagined the two draws from the specified urn as "one draw from each of two distinct urns, whose contents were each determined in the specified manner but were determined independently." Then we should not find a strong preference for this version of gamble BB^{95-100} over gamble AA.

Indeed, suppose subjects believe in Independent Recomposition. In that case, they might easily imagine this new version of gamble BB^{95-100} to have a win probability close to 50%. For although it is possible in their minds that "both urns" contain at least 95 red balls (or that both contain at least 95 blue balls), it is equally possible to them that “one urn contains at least 95 red balls while the other contains at least 95 blue balls.” In other words, their CEs for this version of gamble BB^{95-100} should certainly not be radically larger than their CEs for gamble AA. If such a radical difference in CEs as we found between the original version of gamble BB^{95-100} and gamble AA were still found under this modified version of BB^{95-100}, this would suggest that a belief in Independent Recomposition is not a factor generating our results.

References

Andreoni, James and Michael A. Kuhn (2019), “Is it safe to measure risk preferences? assessing the completeness, predictive validity, and measurement error of various techniques.” Working Paper, URL https://static1.squarespace.com/static/5c79b3d29b8fe82f5cb96360/t/5cc0debb71c10bd5d9ab45f3/1556143804348/mCRB_WP.pdf.

Anscombe, F. J. and R. J. Aumann (1963), “A definition of subjective probability.” Ann. Math. Statist., 34, 199–205, URL https://doi.org/10.1214/aoms/1177704255.

Ellsberg, Daniel (1961), “Risk, ambiguity, and the savage axioms.” The Quarterly Journal of Economics, 75, 643–669, URL http://www.jstor.org/stable/1884324.

Epstein, Larry G and Yoram Halevy (2019), “Ambiguous correlation.” The Review of Economic Studies, 86, 668–693.

Fleurbaey, Marc (2017), “Rationality under risk and uncertainty.” unpublished working paper.

Gajdos, Thibault, Takashi Hayashi, J-M Tallon, and J-C Vergnaud (2008), “Attitude toward imprecise information.” Journal of Economic Theory, 140, 27–65.

Ghirardato, Paolo, Fabio Maccheroni, and Massimo Marinacci (2004), “Differentiating ambiguity and ambiguity attitude.” Journal of Economic Theory, 118, 133 – 173, URL http://www.sciencedirect.com/science/article/pii/S0022053104000262.

Gilboa, Itzhak and David Schmeidler (1989), “Maxmin expected utility with non-unique prior.” Journal of Mathematical Economics, 18, 141 – 153, URL http://www.sciencedirect.com/science/article/pii/0304406889900189.

Gillen, Ben, Erik Snowberg, and Leeat Yariv (2019), “Experimenting with measurement error: Techniques with applications to the caltech cohort study.” Journal of Political Economy, 127, 1826–1863, URL https://doi.org/10.1086/701681.

Halevy, Yoram (2007), “Ellsberg revisited: An experimental study.” Econometrica, 75, 503–536, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2006.00755.x.

Jabarian, Brian (2019), "The moral burden of ambiguity aversion." P&PA Discussion Series, PEA Soup, May 2019, URL http://peasoup.us/2019/05/ppa-discussion-thomas-rowe-and-alex-voorhoeves-egalitarianism-under-severe-uncertainty-with-critical-precis-by-brian-jabarian/.

Jabarian, Brian (2021), "Operationalizing complex uncertainty." Doctoral Thesis in Philosophy, University Panthéon-Sorbonne, Paris 1.

Ju, Nengjiu and Jianjun Miao (2012), "Ambiguity, learning, and asset returns." Econometrica, 80, 559–591.

Karni, Edi and Zvi Safra (1987), "'preference reversal' and the observability of preferences by experimental methods." Econometrica, 55, 675–685, URL http://www.jstor.org/stable/1913606.

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji (2005), "A smooth model of decision making under ambiguity." Econometrica, 73, 1849–1892, URL http://www.jstor.org/stable/3598753.

Kuzmics, Christoph, Brian Rogers, and Xiannong Zhang (2020), "Is ellsberg behavior evidence of ambiguity aversion?" Available at SSRN 3437331.

Lange, Andreas and Nicolas Treich (2008), "Uncertainty, learning and ambiguity in economic models on climate policy: some classical results and new directions." Climatic Change, 89, 7–21.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini (2006), "Ambiguity aversion, robustness, and the variational representation of preferences." Econometrica, 74, 1447–1498, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2006.00716.x.

Millner, Antony, Simon Dietz, and Geoffrey Heal (2013), "Scientific ambiguity and climate policy." Environmental and Resource Economics, 55, 21–46.

Raiffa, Howard (1961), "Risk, ambiguity, and the savage axioms: comment." The Quarterly Journal of Economics, 75, 690–694.

Savage, Leonard J (1954), The foundations of statistics. New York: John Wiley and Sons. Revised and Enlarged Edition, New York: Dover Publications, 1972.

Schmeidler, David (1989), "Subjective probability and expected utility without additivity." Econometrica, 57, 571–587, URL http://www.jstor.org/stable/1911053.

Strzalecki, Tomasz (2011), "Axiomatic foundations of multiplier preferences." Econometrica, 79, 47–73, URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA8155.

Treich, Nicolas (2010), "The value of a statistical life under ambiguity aversion." Journal of Environmental Economics and Management, 59, 15–26.

Yang, Chun-Lei and Lan Yao (2017), "Testing ambiguity theories with a mean-preserving design." Quantitative Economics, 8, 219–238.

Appendix

A Main Tables

A.1 Raw Variable Names

Name	Description	Win Probability
R^j	jth elicitation of CE for 50-50 urn of Ellsberg	.5
A^j	jth elicitation of CE for ambiguous urn of Ellsberg	x
RR^j	jth elicitation of CE for 50-50 urn in 2BallMixed	.5
AA^j	jth elicitation of CE for ambiguous urn in 2BallMixed	$x^2 + (1-x)^2 \geq .5$
AR^j	jth elicitation of CE for "1st urn=A, 2nd=R" gamble of 2BallMixed	.5
RA^j	jth elicitation of CE for "1st urn=R, 2nd=A" gamble of 2BallMixed	.5
$R3$	CE for 3Ball with all three urns = R	.25
$A3$	CE for 3Ball with all three urns = A	$x^3 + (1-x)^3 \geq .25$
RAA	CE for 3Ball with 1st urn = R, latter two urns = A	$.5[x^2 + (1-x)^2] \geq .25$
IA	CE for Independent (Two-Ball gamble with independent ambiguous urns)	$x_1x_2 + (1-x_1)(1-x_2)$†
C^j	jth elicitation of CE for single-urn gamble of Compound	p
CC^j	jth elicitation of CE for Two-Ball gamble of Compound	$p^2 + (1-p)^2 \geq .5$
BB^{40-60}	CE for BoundedA with ambiguous urn containing 40-60 red balls	$x^2 + (1-x)^2 \in [.5, .52]$
BB^{60-100}	CE for BoundedA with ambiguous urn containing 60-100 red balls	$x^2 + (1-x)^2 \geq .52$
BB^{95-100}	CE for BoundedA with ambiguous urn containing 95-100 red balls	$x^2 + (1-x)^2 \geq .905$

Table 5: raw variable names
In the final column, x denotes a number between 0 and 1 that is determined by an ambiguous procedure that is not known by subjects. In reality, x was determined to be one of $0, .01, .02, \cdots, .99, 1$ uniformly at random. x_1 and x_2 denote numbers between 0 and 1 that were determined ambiguously but using the same procedure as each other. Lastly, p is a number between 0 and 1 that subjects know will be determined uniformly at random among $0, .01, .02, \cdots, .99, 1$.
† Note that the win probability for gamble IA will equal .5 if the procedure determining x_1 and x_2 is symmetrical about $.5$ - that is, if the urns are just as likely to contain a certain number of red balls as that to contain that same number of blue balls. Otherwise this win probability will be greater than .5. See the footnote in Section ??.

A.2 Summary Statistics for Raw Variables

	R		*A*		*C*		*RA*	
Mean	118.55	117.72	105.42	106.22	108.67	111.55	95.67	93.72
95% Conf. Interval	[114.43 , 122.66]	[113.48 , 121.96]	[101.10 , 109.74]	[101.81 , 110.63]	[100.15 , 117.19]	[102.31 , 120.79]	[87.64 , 103.70]	[85.54 , 101.89]
ρ	0.875		0.871		0.878		0.855	
	(0.018)		(0.018)		(0.038)		(0.039)	
N	708		708		158		179	
	RR		*AA*		*CC*		*AR*	
Mean	110.55	108.64	100.95	101.12	102.53	104.97	97.21	95.78
95% Conf. Interval	[106.25 , 114.85]	[104.44 , 112.85]	[96.59 , 105.30]	[96.60 , 105.64]	[93.51 , 111.56]	[95.49 , 114.45]	[89.15 , 105.26]	[87.61 , 103.95]
ρ	0.866		0.876		0.855		0.883	
	(0.019)		(0.018)		(0.042)		(0.035)	
N	708		708		158		179	

Table 6: raw variables: decomposed summary statistics

A.3 Summary Statistics for Derived Variables

	$R-A$		$R-C$		$R-AA$		$R-CC$	
Mean	13.15	11.54	7.91	5.70	17.66	16.68	14.11	12.34
95% Conf. Interval	[10.46 ,]	[8.64 , 14.44]	[3.37 , 12.45]	[0.23 , 11.16]	[14.53 , 20.78]	[13.34 , 20.02]	[6.91 , 21.32]	[, 19.82]
ρ	0.489		0.364		0.578		0.561	
	(0.033)		(0.075)		(0.031)		(0.066)	
N	708		158		708		158	
	$R-RR$		$RA-AR$		$RA-AA$		$RA-RR$	
Mean	8.02	9.08	-1.51	-2.07	0.84	-2.85	-11.68	-11.01
95% Conf. Interval	[5.33 ,]	[6.40 , 11.77]	[-4.85 , 1.84]	[-5.48 , 1.35]	[-4.76 , 6.44]	[-7.58 , 1.87]	[-16.81 , -6.55]	[-15.47 , -6.54]
ρ	0.428		0.151		0.368		0.259	
	(0.034)		(0.074)		(0.070)		(0.073)	
N	708		179		179		179	

Table 7: derived variables: decomposed summary statistics

B Which Savage models are refuted by our results?

Our main paper shows how our experimental results falsify any model of decision-making that uses the framework of Anscombe and Aumann (1963) and contains a monotonicity axiom. However, some models of decision-making instead use the framework of Savage (1954), wherein no such concept as "objective probability" exists. Indeed, in the Savage framework, each "state" must encompass how all uncertainty will be resolved. If a decision-maker's preferences over acts satisfy certain properties, the Savage model then defines subjective probabilities that represent that decision-maker's "beliefs" about how likely are the various states - whether or not those subjective probabilities match some "objective" probabilities that one could calculate for those states.

If we allow arbitrary subjective probabilities - i.e. subjective probabilities that have no relationship with the facts of the experiment that are described to the decision-maker (DM) - then there is nothing stopping the DM from believing things such as "A draw from urn R will always be Black, and two consecutive draws from urn A will always be of opposite colors." Such beliefs would be consistent with the axioms of probability theory (and they would induce a preference for gamble R over gamble AA) but they would in no way reflect the realities of the experiment.

Thus, our experimental results are certainly consistent with Savage's theory if we do not introduce any further axioms constraining the DM's preferences over acts to be consistent with the realities of the gambles presented to her. Therefore, we will demonstrate that if we introduce some axioms to minimally constrain the DM's preferences to be consistent with the realities of our gambles, then the preferences exhibited by individuals in our experiment are not consistent with Savage's theory.

Below, we use the colors White (W) and Black (B) for balls in urns, and the letters R and A respectively denote the "risky" (50 White balls, 50 Black balls) and "ambiguous" (unknown proportions of White and Black balls) urns from our experiment.[12]

Our framework is as follows. A state is a tuple (n, r, a_1, a_2) where $n \in \{0, 1, \cdots, 100\}$ and $r, a_1, a_2 \in \{W, B\}$. n represents the number of White balls in urn A, while r represents the color of ball (Black or White) that would be drawn from urn R and a_1 and a_2 respectively represent the 1st and 2nd balls that would be drawn from urn A. We let Ω denote the set of all such states.

We wish to prove that if a DM's preferences satisfy Savage's axioms along with a few axioms that express the fact that "the DM's preferences have to be consistent with the

[12]In our experiment, balls are Red or Black. However, since the letter R denotes the "risky" urn, to avoid confusion about whether and "R" means a color of a ball or a type of urn, here we speak of the color White instead of Red.

information we've given her about our gambles," then she cannot strictly prefer gamble R to gamble AA.

Let "1" denote winning the monetary prize ($3 in our experiment) and "0" denote not winning the monetary prize. Let $\succsim$ be the DM's preferences. As in Savage's framework, let $1_E 0$ denote the act that pays out the monetary prize in states in the event E and pays out nothing otherwise. We assume the following axioms:

A0. $\succsim$ satisfies Savage's axioms, and also $1 \succ 0$ (i.e., the constant act paying out the monetary prize is preferred to the constant act paying out nothing).

By Savage's Theorem, we know that A0 implies that the DM has a subjective probability measure $\mathbb{P}$ on states and a utility function $U : \{0, 1\} \to \mathbb{R}$ such that $U(1) > U(0)$. In our case where there are only the two prizes 1 and 0, we know by Savage's axiom P4 that for any two events A and B,

$$\mathbb{P}[A] \geq P[B] \iff 1_A 0 \succsim 1_B 0. \tag{1}$$

Thus, to show that our DM's preferences must satisfy $AA \succsim R$, it suffices to show that

$$\mathbb{P}\Big[\{(n, r, a_1, a_2) \in \Omega : \text{ act } AA \text{ wins}\}\Big] \geq \mathbb{P}\Big[\{(n, r, a_1, a_2) \in \Omega : \text{ act } R \text{ wins}\}\Big]. \tag{2}$$

To show this, we need to introduce some axioms that specify that the DM's preferences must reflect the information given to her about the gambles.

A1. Let $[R = W]$ denote the event that we draw a White ball from urn R, i.e. $[R = W] = \{(n, r, a_1, a_2) \in \Omega : r = W\}$. Similarly, let $[R = B] = \{(n, r, a_1, a_2) \in \Omega : r = B\}$ be the event that we draw a Black ball from urn R. Then

$$1_{[R=W]}0 \quad \sim \quad 1_{[R=B]}0.$$

By (1), A1 implies that $\mathbb{P}[R = W] = \mathbb{P}[R = B]$, which means that

$$\mathbb{P}\Big[\{(n, r, a_1, a_2) \in \Omega : \text{ act } R \text{ wins}\}\Big] = .5.$$

Thus, to prove (2) and be finished, it suffices to show that

$$\mathbb{P}\Big[\{(n, r, a_1, a_2) \in \Omega : \text{ act } AA \text{ wins}\}\Big] \geq .5. \tag{3}$$

This will follow from our last axiom:

A2. (Some axiom that implies that conditional on the ambiguous urn's number of white balls N, the two draws A_1 and A_2 from urn A are independent of each other and are identically distributed.)

(In fact, these draws are also independent of the draw from urn R, but we don't need this to complete our proof.)

For any $i \in \{0, 1, \cdots, 100\}$, let $[N = i]$ denote the event that urn A contains exactly i white balls, i.e.

$$[N = i] = \{(n, r, a_1, a_2) \in \Omega : n = i\}.$$

Similarly, for any $x \in \{W, B\}$, let $[A_1 = x]$ denote the event that the first ball drawn from urn A will have color x, and let $[A_2 = x]$ be the event that the second ball drawn from urn A has color x.

Given A2, we can argue the following:

$$\mathbb{P}\Big[\{(n, r, a_1, a_2) \in \Omega : \text{ act } AA \text{ wins}\}\Big]$$

$$= \mathbb{P}\Big[\{(n, r, a_1, a_2) : (a_1, a_2) = (W, W) \text{ or } (a_1, a_2) = (B, B)\}\Big]$$

$$= \mathbb{P}\Big[\big([A_1 = W] \cap [A_2 = W]\big) \cup \big([A_1 = B] \cap [A_2 = B]\big)\Big]$$

$$= \mathbb{P}\big([A_1 = W] \cap [A_2 = W]\big) + \mathbb{P}\big([A_1 = B] \cap [A_2 = B]\big)$$

(since these events are disjoint, and $\mathbb{P}$ must satisfy the axioms of probability)

$$= \sum_{i=0}^{100} \Big[\mathbb{P}\big([A_1 = W] \cap [A_2 = W] \mid N = i\big) + \mathbb{P}\big([A_1 = B] \cap [A_2 = B] \mid N = i\big)\Big] \cdot \mathbb{P}[N = i].$$

In this last line, we do not worry about the fact that these conditional probabilities are not defined if the individual's subjective probability $\mathbb{P}[N = i]$ is 0. Indeed, in this case, the term in large brackets (that contains all the conditional probabilities) will be multiplied by $\mathbb{P}[N = i] = 0$ and hence will not contribute anything to the sum. Thus, interpreting the expression in this way, this last line is a legitimate application of the Law of Total Probability.

To proceed from here, we just notice that A2 grants independence between the two draws from urn A once we know condition on the composition of urn A. Thus, we can factor the probabilities:

$$= \sum_{i=0}^{100} \Big[\mathbb{P}(A_1 = W | N = i) \cdot \mathbb{P}(A_2 = W | N = i) + \mathbb{P}(A_1 = B | N = i) \cdot \mathbb{P}(A_2 = B | N = i)\Big] \cdot \mathbb{P}[N = i].$$

Using the fact from A2 that the draws from urn A are conditionally identically distributed, this equals

$$= \sum_{i=0}^{100} \Big[\mathbb{P}(A_1 = W|N = i)^2 + \mathbb{P}(A_1 = B|N = i)^2 \Big] \cdot \mathbb{P}[N = i].$$

Finally, using the fact that $\mathbb{P}$ must satisfy the axioms of probability and that the events $[A_1 = W]$ and $[A_1 = B]$ are mutually exclusive and exhaustive, this equals

$$= \sum_{i=0}^{100} \Big[\mathbb{P}(A_1 = W|N = i)^2 + \big(1 - \mathbb{P}(A_1 = W|N = i)\big)^2 \Big] \cdot \mathbb{P}[N = i].$$

Since the inequality $p^2 + (1-p)^2 \geq .5$ holds for any $p \in [0, 1]$, this implies the inequality

$$\geq \sum_{i=0}^{100} .5 \cdot \mathbb{P}[N = i] = .5,$$

where the last equality follows since the events $[N = 0]$, $[N = 1]$, $\cdots$, $[N = 100]$ are mutually exclusive and exhaustive. Thus, we have show that (3) holds, as desired.

C Future Research: Is it really a distaste for the mere presence of ambiguity?

C.1 New Experiment 1

Description of Gamble. “Urn R has 50 red and 50 blue; urn A has 100 balls in total, all red or blue, with at least 50 of them red. You win if you draw a red ball. Do you prefer to play this gamble with urn R or urn A?” Or, perhaps we guarantee instead that “Urn A contains between 50 and 60 red balls; the rest of its 100 balls are blue.”

Expected finding. People prefer urn A since it has at least as high of a chance of winning as urn R does.

Possible Critique from this finding. People don’t exhibit any distaste for the mere presence of ambiguity; they merely fail to calculate odds correctly when you make things opaque/complicated enough. All of our 2Ellsberg findings are an artifact of the fact that we’ve framed the gambles one way rather than a more straightforward way.

Responses to these critiques. Notice that people do “correctly” identify that $BB^{95-100} \succ BB^{60-100} \succ BB^{40-60}$. Furthermore, their preference for RR over AA is robust to being “nudged” by the BoundedA block. This all suggests that the original preference for RR over AA cannot entirely be due to “a lack of understanding that more unequal urns are better in a 2-ball gamble.”

But what, then, could explain why our results show a distaste for ’ambiguity that can only help you’ while New Experiment 1 shows the opposite? Perhaps the key difference is that New Experiment 1 frames things in a way that immediately suggests a probabilistic dominance of urn A over urn R, while our AA vs. RR question does not. Indeed, perhaps most people do not employ probabilistic thinking in pretty much any scenarios - they only use probabilities when “forced” to do so by the odds of winning being given to them (nearly) explicitly. A comparison between urns A and R in New Experiment 1 forces the observation that “the minimum win probability in urn A is at least as high as the win probability in urn R,” but in 2Ellsberg it does not suggest this observation since the conditional win probabilities (for each ball composition of urn A) are ’hidden’.

C.2 New Experiment 2

Description of Gamble. Elicit CEs for an AA gamble but this time specify that urn A has one of the following three ball compositions:

- 50 red balls and 50 blue balls.
- 75 red balls and 25 blue balls.
- 25 blue balls and 75 red balls.

Also, elicit people's CEs for 2-ball gambles from risky urns (call them urns R, S, and T) that are 50-50, 75-25 and 25-75 in composition. Randomize the order of whether you ask about gambles RR, SS and TT before or after gambles AA and RR.

In each of the 75-25 cases, urn A has a .625 probability of winning. It would be interesting (and a counterexample to Savage, etc.) if people prefer the 75-25 risky urns to the 50-50 risky urn but prefer the 50-50 risky urn A above.

This experiment has the advantage of being simpler than our current experiment - it only has 3 possibilities instead of 101.

We could try also running the same experiment but with e.g. 60-40 and 40-60 in place of 75-25 and 25-75 above. Try also e.g. 90-10 and 10-90. See how extreme you have to make the asymmetry before people exhibit a preference for RR over AA.

C.3 Can People be "Nudged" into Avoiding Dominated Options?

In treatment learning, we exposed subjects to gambles that could help them understand that more uneven urns have higher win probabilities in two-ball gambles (if they did not already understand this). Exposure to these gambles constitutes a very indirect form of learning - subjects were never told that more uneven urns are better; instead they were given a chance to figure this out for themselves if they had not done so already. We found that this indirect learning did not at all reduce subjects' "paradoxical" choice of the dominated gamble R over the ambiguous gamble AA. We therefore concluded that subjects' choice to avoid AA is not due to a lack of understanding but instead due to a distaste for the presence of ambiguity.

In contrast, Kuzmics et al. (2020) found that subjects' paradoxical choice for avoiding draws from ambiguous urns - even at the cost of choosing a dominated option - can be reduced by providing information that clarifies how a certain option, potentially involving ambiguous draws, yields a larger win probability than the unambiguous option. Specifically, in two of their experimental treatments they show subjects two videos, both containing factually correct information, prior to eliciting subjects' choices. One video, "V1," argues why a Raiffa (1961)-style choice to bet on a single draw from an ambiguous urn, choosing the color on which to bet based on the result of a coin flip, will increase the probability of winning relative to an unambiguous 49% win probability gamble. Meanwhile, the other video "V2" merely argues that, given that the subject has already bet on a particular color in an ambiguous urn, no conclusion can be reached about the subject's probability of winning. V1 is meant to provide information that might encourage a choice of a "better" option (the Raiffa-style one that involves drawing from an ambiguous urn), while V2 is meant to provide information that might encourage avoiding options involving ambiguous draws.

The authors include these videos in parts of two treatments. In their "coin" treatment the authors allow subjects to commit to placing a bet on the ambiguous urn based on the result of a coin flip carried out for them automatically, while in their "no coin" treatment they do not offer this option but merely suggest that subjects could imagine flipping a coin for themselves. In parts of both treatments they show subjects videos V1 and V2 before eliciting choices; in some other parts they do not show these videos. In the "coin" treatment they find that exposure to V1 and V2 decreases the proportion of subjects who choose a dominated option (that involves no ambiguity), while in the "no coin" treatment they find that such exposure increases it. The authors therefore argue that subjects' choices to avoid options involving ambiguity - even if it means choosing dominated options - is not due to a deliberate preference but instead due to a lack of understand of the options before them.

What might explain the difference in results between our experiment and that of these authors? One possible explanation is that subjects do (at least mostly) understand the options before them in both experiments and that videos V1 and V2 mostly create an "experimenter demand" rather than additional understanding - with the experimenter demand for the Raiffa option in V1 being stronger than the experimenter demand for the unambiguous option in V2. Indeed, in the "coin" treatment subjects can explicitly demonstrate compliance with the experimenters' suggestions by having their choice to bet using the Raiffa coin toss be recorded as such, while in the "no coin" treatment they have no such option and instead opt to record themselves satisfying the (weaker) suggestion of V2 to avoid ambiguous draws. In contrast, our experiment does not make any explicit arguments suggesting why subjects might want to choose one option or another; it merely presents them with choice problems that can help create understanding if it does not exist already. Such learning induces no change in behavior since it does not create experimenter demand.

It is also possible that this disparity between our results and these authors' results is simply due to a difference in the nature of the experiments: perhaps subjects generally understand the gambles in our experiment without the need of any explanation, while the same is not true of Kuzmics et al. (2020)'s experiment. In this case, further research is warranted to determine the difference between those circumstances in which subjects can be "nudeged" into choosing dominant-but-ambiguous options and those in which they cannot.

Online Appendix

D Variable Names

Name	Definition	Description
E^j	$K^j - U^j$	CE difference in j-th elicitation of Ellsberg
Z^j	$K^j - UU^j$	CE difference in j-th elicitation of 2Stage
H^j	$K^j - C^j$	CE difference in j-th elicitation of 50-50 vs. Halevy compound 50-50
L^j	$K^j - CC^j$	CE difference in j-th elicitation of 2-Stage simple 50-50 vs. compound 50-50
I^E	$\mathbb{1}\{E^1 + E^2 > 0\}$	Indicator for falling for classic Ellsberg paradox
I^T	$\mathbb{1}\{T^1 + T^2 > 0\}$	Indicator for falling for 2-Stage Ellsberg paradox
I^H	$\mathbb{1}\{H^1 + H^2 > 0\}$	Indicator for falling for Halevy paradox
I^L	$\mathbb{1}\{L^1 + L^2 > 0\}$	Indicator for falling for unambiguous 2-Stage paradox
F^{0-2}	$.5(UU^1/KK^1) + .5(UU^2/KK^2)$	Ratio of certainty equivalents for UU and KK (averaged across 2 elicitations)
F^{0-3}	$UUU/3K$	Ratio of certainty equivalents for UUU and KKK
F^{1-3}	$KUU/3K$	Ratio of certainty equivalents for KUU and KKK
I^B	Treatment = $\mathcal{C}$ & did "Bounded U" first	Indicator variable for having the "learning" section first
I^R	$R^2 = 1$ and $R^3 = 0$	Indicator variable for choosing the correct color in both practice questions
I^A	all $A^j = 1$	Indicator variable for get all 3 attention screeners correct

Table 8: Contingent Variable Names

	T^1	T^2	E^1	E^2	H^1	H^2	L^1
Mean	9.92	6.90	13.49	10.79	1.21	6.74	11.31
95% Conf. Interval	[6.39 , 13.44]	[3.43 , 10.38]	[9.99 , 17.00]	[7.22 , 14.36]	[-7.10 , 9.52]	[-1.28 , 14.76]	[2.73 , 19.89]
ρ	0.152		0.287		0.134		0.22
$1-\rho$	0.848		0.713		0.866		0.77
se $1-\rho$	0.033		0.032		0.067		0.06
N	880		880		220		220

Table 9: Decomposed Summary Statistics

E Experimental Design Details

E.1 Treatments Blocks and Gambles

Our experiment contains four treatments, each comprising a specific number of blocks of gambles. A block contains either one or several similar gambles. Before each block, subjects view the relevant instructions. Each elicitation within a given block contains (1) a reiteration of the block's instructions, (2) the new details of that particular elicitation, highlighted in yellow, and (3) a report of the subjects' CE for that elicitation.[13] Subjects must report their CE before moving on to the next elicitation screen. Elicitations are uniformly, independently, and randomly ordered between the subjects within a given block. Each treatment may only contain 11 or 12 elicitations to accommodate online cognitive fatigue and prevent attention deficits.

Each treatment is divided into blocks consisting of one or multiple questions about a gamble for which the subjects must report their CEs.

In each question, "winning" the gamble means a payoff of 300 tokens (=$3), and "losing" means a payoff of 0 tokens. The notation "[x red, y blue]" means an urn that contains exactly x red balls, y blue balls, and no other balls. Similarly, "[Unknown red, Unknown blue]" means the urn contains an unknown number of red and blue balls and no other balls. For notational convenience, = [50 red, 50 blue] and A= [Unknown red, Unknown blue].

Subjects were informed that the contents of urn A would vary from question[14] to question (i.e., the contents of ambiguous urns are re-determined between questions). In practice, the contents of each urn A were determined by drawing an integer X uniformly at random between 0 and 100. A virtual urn containing X red balls and $100 - X$ blue balls was created. Subjects were not informed of this procedure to determine the contents of ambiguous urns.

To perform ORIV, we double-elicit subjects' CEs for all gambles of central importance to our analysis; however, due to time constraints and concerns that subjects may "zone out" and provide especially noisy answers if asked too many repeated similar questions, we could not double-elicit CEs for all gambles. We focused on double eliciting the most relevant gambles to our paper. We will attach the symbol D to the name of an elicitation when we refer to a duplicate of this later.

Table 10 summarizes the structure of each treatment. Each item in bold is one of

[13] Certain elicitations require the subject to choose a color (i.e., red or blue) to place a bet. For these elicitations, the subject must select a color before they can report their CE, which appears on the screen.

[14] In the remaining of our paper, we will use the words "elicitations" and "questions" as synonyms.

the blocks described in Section E.1. Multiple items within parenthesis ( ) mean that the order of these items is determined uniformly at random, independently for each subject. Items within brackets [ ] are not randomized; they always appear in the order listed within them.

In each treatment, we double-elicit subjects' CEs for the two classic Ellsberg gambles as well as the two Two-Ball gambles in the 2Ball block (which also appear within its longer version 2BallMixed). Thus, using data from all four treatments, we can robustly determine if subjects prefer RR (or R) over AA, even though the latter is more likely to win. Furthermore, by comparing a subject's responses to these Two-Ball gambles with their responses to the classic Ellsberg gambles, we can determine the relationship between ambiguity aversion, risk aversion, and "falling for" the Two-Ball Ellsberg paradox.

Treatment	Contents of Treatment
paradoxes	[(Ellsberg, 2BallMixed), (EllsbergD, 2BallMixedD)]
complexity	[(Ellsberg, 2Ball, Compound), (EllsbergD, 2BallD, CompoundD)]
nudging	(BoundedA, [(Ellsberg, 2Ball), (EllsbergD, 2BallD)])
robustness	((Ellsberg, 2Ball), (3Ball, Independent), (EllsbergD, 2BallD))

Table 10: treatments

E.2 Elicitation Protocol: Multiple Price List

As mentioned in the introduction, we elicit the subjects' CEs using MPLs to determine their preferences over various acts. Each question introduces a gamble, as detailed above. When agents do not make choices that correspond to the expected utility theory predictions, using the MPL mechanism may be problematic. For example, Karni and Safra (1987) demonstrated that incentive-compatible mechanisms could not elicit CEs if the independence axiom does not hold. Despite this concern, the MPL mechanism has been used extensively in experiments where agents face risk or ambiguity when making choices, many of which included the possibility of their choices over lotteries not satisfying the predictions of expected utility theory. This is perhaps because the MPL offers several advantages over other mechanisms. Andreoni and Kuhn (2019) argue that the MPL mechanism is extremely easy for subjects to understand and yields more consistent choices than other standard mechanisms for eliciting risk preferences. Furthermore, it provides externally valid predictions once adjusted for measurement error.

Our experiment's MPL table contains 31 rows corresponding to fixed prize values between 0 and 300 tokens in increments of 10 tokens. There are 32 possible locations where a subject can place their "cutoff" (below which they prefer the gamble and after which they prefer the fixed prize). If a value $x \in \{0, 10, \dots, 290\}$ exists such that the subject prefers the gamble to receive x tokens but prefers receiving $x + 10$ tokens to the gamble, then this was recorded numerically as "the subject's CE is $x + 5$." If the subject preferred 0 tokens to the gamble, the CE was 0. Finally, if the subject preferred the gamble to 300 tokens, the CE was 300.

In each row, subjects select either the left column ("Receive fixed payment") or the right column ("Play the gamble"). To make the process less time-consuming and enforce the consistency of choices, the subject's selection in each row is automatically completed based on a limited number of clicks. For example, suppose a subject clicks to indicate a preference for 150 tokens instead of the gamble. In that case, the JavaScript algorithm automatically completes rows 160 through 300 to indicate that the subject prefers receiving tokens to the gamble. Similarly, if the subject prefers the gamble instead of receiving 140 tokens, the software automatically completes rows 0 through 130 to indicate a preference for playing the gamble over receiving tokens. Subjects can revise their choices (consistent with the autocompletion rules above) before moving on to the next question.

Each question contains, at most, one row in which the subject's preference switches from preferring the gamble to preferring a specific amount of tokens. The subject's CE for the gamble must lie between the token amounts in this row and the previous row. We then record the subject's CE as the midpoint between the two rows, i.e., a number ending in 5. If the subject prefers the gamble over 300 tokens or 0 tokens to the gamble, then no such "switching" row exists. Nonetheless, if the subject prefers the gamble over a fixed payment of 300 tokens, their CE may be 300 tokens, as the gamble cannot pay more than 300 tokens. Similarly, if the subject prefers 0 tokens to the gamble, their CE is 0. We record the subject's CE as 300 or 0 in these cases.

E.3 Payment Method: Fixed Sum and Incentive Mechanism

Fourteen questions are selected uniformly at random for payment from among all the questions in a given treatment to make this mechanism incentive compatible. Some experiments eliciting risk attitudes select only a single question for payment, avoiding the possibility of subjects using their choices in different questions to hedge their payoffs; however, doing so creates a significant variance in the monetary payments that different subjects receive, which was undesirable for this experiment. If a question is selected for payment, then one row of that question's MPL table is selected randomly, and the subject is given whatever their preference is in that row. For example, if row 120 was selected and the subject preferred the gamble to 120 tokens, then the gamble is simulated, and the subject wins the prize (usually 300 tokens) or receives 0 tokens if they lose. If the subject preferred 120 tokens to the gamble, they would receive 120 tokens.

To eliminate the possibility of wealth effects and ensure that subjects did not "learn" the distribution used to resolve ambiguity, the payoffs for each question (as well as which questions were selected for payment) were not determined until after the subject completed the entire experiment. Subjects were invited to practice with the MPL mechanism (before the experiment) and observe a summary of the results; they were informed that these practice questions would not be selected for payment. Furthermore, none of these questions involved ambiguity; hence, none presented an opportunity to learn how this experiment resolved ambiguity.

At the end of the experiment, subjects were presented with a table summarizing the questions selected for payment, the row selected in that question's MPL, the subject preference in that row, and (if they preferred the gamble) whether they won the gamble. Moreover, the subject's total payment was $1 for every 100 tokens earned, in addition to a fixed payment of $2 for participation.

E.4 Double Elicitations, Measurement Error, and Attention Screeners

As mentioned in the introduction, laboratory experiments eliciting subjects' CEs for gambles are often subject to significant measurement errors. Such errors can create significant bias in estimated correlations and regression coefficients if not considered. Methods to correct for such measurement error involve eliciting subjects' CEs twice for each gamble of interest.

Although many techniques can then be used to eliminate the bias in estimating coefficients and correlations; the ORIV proposed in Gillen et al. (2019) generally estimates these parameters with lower standard errors. Hence, we rely on the latter. Essentially, this estimation entails using multiple instrumentation strategies simultaneously, then combining the results.

Due to the complex nature of some of the questions, it is concerning that some subjects may not comprehend the questions or may give random responses to complete the experiment quickly. Although most of the financial reward comes from incentivized MPL questions, there is a small fixed reward for merely completing the experiment. To avoid this concern, subjects were screened based on three criteria:

(1) After receiving general instructions concerning the experiment, subjects were given a basic comprehension quiz with three questions regarding those instructions. Subjects unable to correctly answer the three questions were removed from the experiment. They received a small fixed amount for their two-minute participation and were made aware of this scenario when they offered their consent.

(2) Between each of the experiment's major sections, subjects were given a standard attention-screening question.

(3) If, in the course of our double elicitation of a subject's preferences, two reported CEs for the same question differed by more than 100 tokens—that is, one-third the size of the 300-token table—then the subject was deemed to be paying insufficient attention to the experiment.[15]

Subjects failing criterion (1) were immediately removed from the experiment and received a minimum payment.[16] Subjects failing at least one of the attention-screening questions in (2) were subsequently removed. Finally, subjects deemed to be paying insufficient attention were removed according to (3). As a result, out of an initial 880 subjects, 172 were excluded from our data set.

F Prolific Data Collection Details

F.1 Fair Attention Check

We used attention checks. This has been developing these last few years. However, amid those attention checks, some are valid and others are not. Those not valid are..;Those valid, called "fair attention checks" are...We used these latter ones, following Prolific standards.

[15] Other thresholds for exclusion, such as "differed by more than 150 tokens," yield qualitatively similar results to those below. See Appendix.

[16] See section ?? for details.

F.2 Preventing Duplicates

Submissions to studies on Prolific are guaranteed to be unique by the firm[17]. Our system is set up such that each participant can have only one submission per study on Prolific. Each participant will be listed in your dashboard only once, and can only be paid once. On our side, we also prevent participants from taking up our experiment several times in two steps. First, we enable the functionality "Prevent Ballot Box Stuffing," which permits to...Second, we check the participant ID and delete the second submission from the data set of the same ID if we find any.

Drop-out Rates. Here, put the drop out (or in the main text).

F.3 High vs. Low-quality Submissions

Participants joining the Prolific pool receive a rate based on the quality of their engagement with the studies. If they are rejected from a study, then they receive a malus. If they receive too much malus, then they are removed by the pool from the company[18]. Based on this long term contract, participants are incentivized to pay attention and follow the expectations of each study. Hence, a good research behavior has emerged on Prolific according to which participants themselves can vol voluntarily withdraw their submissions if they feel they did a mistake such as rushing too much, letting the survey open for a long period without engaging with it, and so on[19]. According to these standards, we kept submissions rejections as low as possible, following standard in online experimental economics. Participants who fail at least one fair attention check are rejected and not paid. Following Prolific standards, participants who are statistical outliers (3 standard deviations below the mean) are excluded from the good complete data set.

F.4 Payments And Communication

We make sure to review participants' submissions within 24-48 hours after they have completed the study. If we accept their submission, they receive their fixed and bonus payment within this time frame. Otherwise, we reject their submissions and send them a personalized e-mail([20]), detailing the reason for the rejection, leaving participants the

[17] See Prolific unique submission guarantee policy here.

[18] See Prolific pool removal Policy here.

[19] See Prolific update regarding this behavior here.

[20] Partially-anonymized through Prolific messaging app which puts the researcher's name visible to the participants and only the participants visible to the researcher.

opportunity to contact us afterward if they firmly believe the decision to be unfair (motivate their perspective). Participants can also contact us at any time if they encounter problems with our study or have questions about it.

G Variables Dictionary

G.1 Independent Variables

Stata/Paper	Data File	Elicitation Description
K^1	Balc1a	1st elicitation of risk preferences in one-stage Ellsberg
K^2	Final1a	2nd elicitation of risk preferences in one-stage Ellsberg
U^1	Balc1d	1st elicitation of ambiguous preferences in one-stage Ellsberg
U^2	Final1c	2nd elicitation of ambiguous preferences in one-stage Ellsberg
KK^1	Balu1a	1st elicitation of risk preferences in two-stage Ellsberg
KK^2	Matu1a	2nd elicitation of risk preferences in two-stage Ellsberg
UU^1	Balu1b	1st elicitation of ambiguous preferences in two-stage Ellsberg
UU^2	Matu1b	2nd elicitation of ambiguous preferences in two-stage Ellsberg
UK^1	Balu1c	
UK^2	Matu1c	
KU^1	Balu1d	
KU^2	Matu1d	
KKK	Balu2a	elicitation of risk preferences in 3-stage Ellsberg
UUU	Balu2b	elicitation of ambiguous preferences in 3-stage Ellsberg
KUU	Balu2c	
II	Balu4	2-Stage gamble with indepedent ambiguous urns
C^1	Lotte1	1st Halevy compound 50-50 lottery
C^2	Final2a	
CC^1	Lotte2	1st 2-stage Halevy
CC^2	Final2b	
BB^{40-60}	Cmu1b	2-stage Ellsberg with bounded U ($40 \leq R \leq 60$)
BB^{60-100}	Cmu2b	2-stage Ellsberg with bounded U ($60 \leq R \leq 100$)
BB^{95-100}	Cmu4b	2-stage Ellsberg with bounded U ($95 \leq R \leq 100$)
R^1	Answered “red” on Mp1	
R^2	Answered “red” on Mp2	Picked the CORRECT color in practice question 2
R^3	Answered “red” on Mp3	Picked the WRONG color in practice question 3
P^1	$Q78$	Indicator variable for get $P^1 = 1$, i.e., correct := “32 Blue balls and 95 Red balls”
P^2	$Q1777$	Indicator variable for get $P^2 = 1$, i.e., correct := “2”
P^3	$Q80$	Indicator variable for get $P^3 = 1$, i.e., correct := “\$1”
A^1	$Q13$	Indicator variable for get $A^1 = 1$, i.e., correct := “orange”
A^2	$Q22$	Indicator variable for get $A^2 = 1$, i.e., correct := “11”
A^3	$Q30$	Indicator variable for get $A^3 = 1$, i.e., correct := “blue”

independent variable names

G.2 Dependent Variables

Note on the naming convention for first few items: E=Ellsberg, T=Two-stage, H=Halevy, L = compound Lottery

Stata/Paper	Definition	Description
E^j	$K^j - U^j$	Certain equivalent difference in j-th elicitation of 1-stage Ellsberg
T^j	$KK^j - UU^j$	Certain equivalent difference in j-th elicitation of 2-stage Ellsberg
H^j	$K^j - C^j$	Certain equivalent difference in j-th elicitation of 50-50 vs. Halevy compound 50-50
L^j	$KK^j - CC^j$	Certain equivalent difference in j-th elicitation of KK vs. CC
F^{0-2}	$.5(UU^1/KK^1) + .5(UU^2/KK^2)$	Ratio of certainty equivalents for UU and KK (averaged across 2 elicitations)
F^{0-3}	UUU/KKK	Ratio of certainty equivalents for UUU and KKK
F^{1-3}	KUU/KKK	Ratio of certainty equivalents for KUU and KKK
I^E	$E^1 + E^2 > 0$	Indicator variable for having a larger Certain equivalent for K than U
I^T	$T^1 + T^2 > 0$	Indicator variable for having a larger Certain equivalent for KK than UU
I^H	$H^1 + H^2 > 0$	Indicator variable for having a larger Certain equivalent for K than C
I^L	$L^1 + L^2 > 0$	Indicator variable for having a larger Certain equivalent for KK than CC
I^B	Treatment $= \mathcal{C}$ & did "Bounded U" first	Indicator variable for having the "learning" section first
I^R	$R^2 = 1$ and $R^3 = 0$	Indicator variable for choosing the correct color in both practice questions
I^A	all $A^j = 1$	Indicator variable for get all 3 attention screeners correct

dependent variable names

MPL Example

Complete experimental instructions available online

In this section, you will be presented with an urn. The gamble is as follows: you get to choose a color (either red or blue), then we will draw a ball at random from the urn. You win 300 tokens if the ball we drew was the COLOR YOU CHOSE.

Suppose the urn is [25 Red, 25 Blue]. Which do you prefer?

RED BLUE

	Receive fixed payment	Play the gamble
Fixed payment: 0 tokens		
Fixed payment: 10 tokens		
Fixed payment: 20 tokens		
Fixed payment: 30 tokens		
Fixed payment: 40 tokens		
Fixed payment: 50 tokens		
Fixed payment: 60 tokens		
Fixed payment: 70 tokens		
Fixed payment: 80 tokens		
Fixed payment: 90 tokens		
Fixed payment: 100 tokens		
Fixed payment: 110 tokens		
Fixed payment: 120 tokens		
Fixed payment: 130 tokens		
Fixed payment: 140 tokens		
Fixed payment: 150 tokens		
Fixed payment: 160 tokens		
Fixed payment: 170 tokens		
Fixed payment: 180 tokens		
Fixed payment: 190 tokens		
Fixed payment: 200 tokens		
Fixed payment: 210 tokens		
Fixed payment: 220 tokens		
Fixed payment: 230 tokens		
Fixed payment: 240 tokens		
Fixed payment: 250 tokens		
Fixed payment: 260 tokens		
Fixed payment: 270 tokens		
Fixed payment: 280 tokens		
Fixed payment: 290 tokens		
Fixed payment: 300 tokens		