Ellsberg's Hidden Paradox

Sean Crockett[*], Yehuda Izhakian[†], and Julian Jamison[‡§]

July 20, 2019

Abstract

Ellsberg (1961) proposes two alternative frames to elicit individuals' preferences for ambiguity. Through an experiment, we find that Ellsberg's three-color one-urn frame induces very different revealed preferences than the two-color two-urn frame. In both frames, we document ambiguity aversion for likely gains and (weak) ambiguity seeking for unlikely gains. The intensity of both these attitudes, however, is strongly curbed in the three-color frame. These findings may imply that perceived probabilities are biased toward uniformly distributed states *and* consequences, introducing a tension between them in the three-color frame. Surprisingly, we find that subjects who are simply asked to fill urns with colored bingo chips for a fixed payment also reveal a similar uniformity bias. We argue that both urn fillers and incentivized decision makers are influenced by similar perceptually salient features of probability intervals. Our findings have important implications for decision theory, and, by extension, for economic, financial, political, and medical decisions.

Keywords: Perceived probabilities, ambiguity, Knightian uncertainty, experiment, framing, judgment and decision making

JEL Classification: C91, D81, D91

[*]Baruch College and University of Melbourne; seanmcrockett@gmail.com

[†]Baruch College; yudizh@gmail.com

[‡]University of Exeter; julison@gmail.com

[§]We acknowledge outstanding research assistance from Florian Mudekereza, Andre Mouton, Gilbert Nyantakyi, Akhtar Shah, Roman Shelkhov, and Nathaniel Zinda. We appreciate helpful comments and suggestions from Peter Bossaerts, Felix Fattinger, Scott Huettel, Edi Karni, David Kelsey, Dan Levin, Carsten Murawski, and Marciano Siniscalchi. We are especially grateful to Peter Wakker for his insightful comments and interpretations. We also thank Michael Donovan and Eileen Ippolito for their assistance with recruiting the participation of fourth grade students as part of this research project.

I Introduction

Ellsberg (1961) suggests that individuals prefer *risk*—an unknown realized event with known probability—to *ambiguity* (*Knightian uncertainty*)—an unknown realized event with unknown probability. He proposes two alternative frames to elicit preferences for ambiguity: a three-color one-urn frame and a two-color two-urn frame. We show that equivalent two-color two-urn representations of three-color one-urn decision problems induce very different revealed preferences. We then attempt to identify the source of this inconsistency.

We present evidence that perception of ambiguity, and therefore perceived probabilities, are biased toward uniformly distributed *states* and *consequences*. Suppose a state space consists of two states. Each state is associated with a different payoff, and its probability is unknown but within a given known range. In this case, most individuals perceive the probability of the favorable payoff to be between the midpoint of the probability range and 50%, leading to choices consistent with ambiguity aversion for likely gains and ambiguity seeking for unlikely gains. This pattern is known as *a-sensitivity* (Dimmock et al., 2016), and shares features with the probability weighting schema (Tversky and Kahneman, 1992; Gonzalez and Wu, 1999). However, when the state space consists of three states, partitioned by two payoffs into two events (so that one of the payoffs follows from two different states), a tension is introduced between bias toward uniformly distributed consequences and uniformly distributed states. The former shifts the perceived probability of the favorable payoff to be between the midpoint of the probability range and 50%, while the latter shifts it to be between the midpoint and 33.3%. In Ellsberg's classic three-color experiment, this tension mitigates perceived probability biases and induces behavior consistent with less intense ambiguity aversion/seeking. This new finding is an ambiguity analog to uniform-biased partition-dependent probability weighting functions (Fox and Clemen, 2005).

Further, this bias toward uniformity does not merely reflect subjective beliefs or attitudes, but appears to be rooted in perception. That is, it appears to be an autonomous rather than deliberative response, consistent with neuroscientific evidence of the impulsivity of choices under ambiguity (Huettel et al., 2006). Supporting this claim, some subjects in our experiment are simply tasked to fill urns with colored bingo chips from a set of allowable distributions, with no induced incentives over the resulting choices. These subjects reveal a similar strong bias toward uniformity.

We deliver our core results through simple but important modifications to the seminal design of

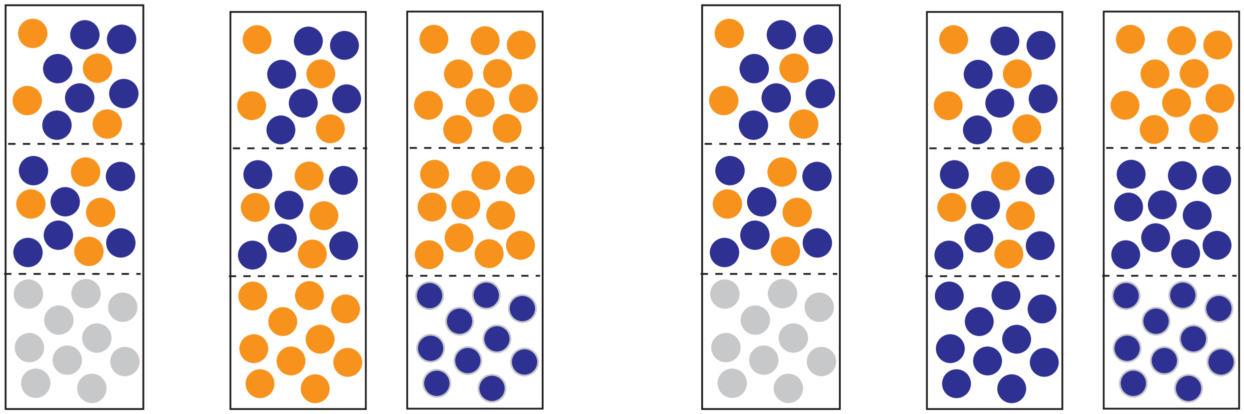
Ellsberg (1961). In Ellsberg's three-color experiment, the decision maker (DM) is presented with an urn that contains blue, orange and gray colored balls.[1] One-third of the balls are gray, while the others are either blue or orange in an unknown proportion. A ball is drawn from the urn at random, and a prize is associated with a correct bet. The experiment consists of two parts. First, the DM must choose between two bets: the next drawn ball is gray (G), or the next drawn ball is blue (B). Then, she must choose between betting that the next drawn ball is gray or orange (GO) or, alternatively, that the next drawn ball is blue or orange (BO). We label these two decision problems as E.I and E.II, respectively.

Ellsberg conjectures that individuals will typically avoid ambiguity in both decision problems, preferring G to B in E.I and BO to GO in E.II. Early experimental support for this conjecture is summarized in Camerer and Weber (1992), although more recent studies identify the revealed aversion to ambiguity in E.I as weak or neutral (e.g., Binmore et al., 2012; Charness et al., 2013; Stahl, 2014). Ambiguity aversion for unlikely gains in E.I presents a puzzle: while ambiguity aversion is widely reported in the literature, a growing number of laboratory studies report ambiguity seeking for unlikely gains (e.g., Mukhtar, 1977; Curley and Yates, 1985, 1989; Di Mauro and Maffioletti, 2004; Abdellaoui et al., 2011; Armantier and Treich, 2015; Dimmock et al., 2016; Kocher et al., 2018), as does Brenner and Izhakian (2018), which extracts attitude toward ambiguity using stock market prices. The review of Trautmann and van de Kuilen (2015) summarizes this literature.

To address this puzzle, we design an equivalent representation of the three-color experiment using two urns and two colors. In decision problem D.I (the "dual" to E.I), urn 1 contains between zero to two-thirds blue balls and the rest are orange, while urn 2 contains exactly one-third blue balls and two-thirds orange balls. A ball is drawn from one of the urns at random, and the winning color is blue (B). The DM must choose from which urn the ball is drawn: from urn 1 or from urn 2. In decision problem D.II (the "dual" to E.II), urn 1 contains between one-third to three-thirds blue balls and the rest of the balls are orange, while urn 2 contains exactly two-thirds blue balls and one-third orange balls. Again, the winning color is blue (B) and the DM must choose from which urn the ball is drawn: from urn 1 or from urn 2. Figures 1a and 1b present a visualization of decision problems E.I and D.I, respectively. Figures 1c and 1d present a visualization of decision problems E.II and D.II, respectively.[2]

[1] In his original experiment, Ellsberg (1961) uses the colors red, black and yellow.

[2] In the experiment, we use colored bingo chips in place of balls, and the winning prize is \$20. In some sessions,



(a) E.I: B or G (b) D.I: Urn 1 or 2 (B wins) (c) E.II: GO or BO (d) D.II: Urn 1 or 2 (B wins)

Figure 1: **Ellsberg three-color experiment and its dual two-urn representation**

Table 1 summarizes the choices made by 102 subjects in these four decision problems in our baseline experiment. The two-color two-urn frame strongly supports a-sensitivity: the majority, 73%, of DMs prefer the *ambiguous* urn in D.I, while a mere 15% of the same DMs prefer the ambiguous urn in D.II. But when the one-to-one relationship between states and consequences is broken in the three-color frame, only 50% of DMs prefer the ambiguous outcome in E.I while 33% prefer the ambiguous outcome in E.II. That is, the presence of the additional state dramatically narrows the strength of revealed preferences, since uniformity bias over states conflicts with uniformity bias over consequences.

Table 1: **Revealed preferences for ambiguity**

	Ambiguous	Risk-only	Prob. Ambiguous	Prob. Risky	Prefer Ambiguous
E.I. Three-color:	B	G	$\left[0, \frac{2}{3}\right]$	$\frac{1}{3}$	50%
E.II. Three-color:	GO	BO	$\left[\frac{1}{3}, 1\right]$	$\frac{2}{3}$	33%
D.I. Two-urn:	Urn 1	Urn 2	$\left[0, \frac{2}{3}\right]$	$\frac{1}{3}$	73%
D.II. Two-urn:	Urn 1	Urn 2	$\left[\frac{1}{3}, 1\right]$	$\frac{2}{3}$	15%

We establish the robustness of our results through additional binary choices with varied ranges of winning probabilities. Further, in some decision problems we elicit the "certainty equivalent probability" (CEP, Izhakian, 2017)—the minimum certain probability value that the DM is willing to accept in exchange for the uncertain probability of a favorable outcome—known also as the "matching probability" (Dimmock et al., 2016). To this end, we implement two popular elicitation

we switch the roles of blue and orange and/or reverse the order of E and D to control for color and order effects.

mechanisms: Becker-DeGroot-Marschak (BDM) mechanism (Becker, DeGroot and Marschak, 1964) and Multiple Price List (MPL).

The experimental literature, surveyed in Trautmann and van de Kuilen (2015), has adopted three alternative methods to determine the composition of ambiguous urns: (i) a data generating process unknown to subjects; (ii) a uniform random lottery implemented after decisions have been recorded; and (iii) natural sources—a non-mathematically characterized data generating process that is revealed to subjects prior to choices (e.g., in Dhillon and Mertens, 1999, based on temperature in some distant city). In some treatments, we adopt source (iii): a human subject who is randomly selected at the beginning of the session to fill the urns later used by other subjects in the experiment. The urn filler, who is given a flat payment for her assistance, is provided with public rules regarding how to fill various urns, but is given no further information about the experiment; i.e., the subject is not informed that other subjects will place bets on these urns or that different colors have different values. In a control treatment, we adopt source (ii) and find that there is no difference in the distribution of choices between sources (ii) and (iii). This crucial finding indicates that *subjects in our experiment do not attempt to predict a bias in the urn-filling choices of other subjects*. A similar source-independent preference for ambiguity is also reported in prior studies (e.g., Halevy, 2007; Oechssler and Roomets, 2015; Armantier and Treich, 2015).

Surprisingly, and central to our contribution, we find that urn-filling subjects reveal strong preferences. We ask subjects to fill two-color urns with N total bingo chips. When subjects are asked to place *at most* b chips of one color and the rest of the other color, they tend to place a number of chips of the first color *greater* than the midpoint, m, of the allowable range. When they are asked to place *at least* b chips of one color and the rest of the other color, they tend to place a number of chips of the first color *less* than m, the midpoint of the allowable range. That is, there is a strong bias toward b, and thus toward uniformity of the two colors for urn-filling choices—the same bias subjects exhibit in incentivized choices but in exaggerated fashion.

In the three-color Ellsberg urn E, urn-filling subjects are instructed to place 9 bingo chips in total: 3 gray chips, 0-6 blue chips, and the rest, 0-6, orange chips. In sharp contrast to all two-color urns, the internal boundary $b = 6$ does not appear to be focal for blue or orange urn-filling choices. Instead, 54% of the urn-filling subjects place exactly three chips—the midpoint of allowable range—of each color in the urn, and 78% choose to place 2-4 blue chips and 2-4 orange chips in the urn.

The urn-filling choices for E are highly symmetric about the midpoint of the range of allowable blue and orange chips, in stark contrast to all two-color urns in the experiment, in which there is a strong bias away from the midpoint of allowable range toward the internal boundary b. Note that urn-filling subjects are not aware of consequences, only states. Thus, uniformity over color choices (states) in filling urn E is particularly strong. These findings suggest that a break in the perceptual salience of the internal boundary in the three-color urn may also *cause* the weakening of revealed preferences in E.I and E.II relative to their respective duals D.I and D.II.

To address the question of whether the pattern we observe is related to the age and maturity of the subjects, or to the inferred financial consequences of their choices on others, we replicate the urn-filling part of the experiment with fourth-grade students (i.e., children aged 9-10 years old). These subjects exhibit a similar pattern of choices as our pool of college student subjects. Thus, although the pattern of urn-filling choices and incentivized bets are aligned, it does not appear that the urn fillers attempt to aid the bettor, as it is unlikely that children could anticipate how the urns they fill will be used. Nor do the bettors appear to leverage beliefs about the choices of urn fillers, since choices when "ambiguity" is generated by a uniform random device are indistinguishable from choices related to subject-filled urns. These findings suggest a perceptual bias in decision making shared by both urn fillers and incentivized decision makers.

Perceptual salience for a subset of alternatives is well documented in many fields. Wertheimer (1938) suggests there are "ideal types" among perceptual stimuli that act as anchoring points. Expanding on this suggestion, Rosch (1975) reports experimental findings of natural reference points in three domains (colors, line orientation, and numbers), which she characterizes as *cognitive reference points*. She writes that, "The underlying hypothesis is that stimuli slightly deviant from reference stimuli are more easily assimilated to and, thus, judged metaphorically closer to the reference stimuli than vice versa" (page 533, Rosch, 1975). Li and Camerer (2019) examine the impact of visual salience on strategic behavior in online hide-and-go-seek games. Closer to our findings, Holyoak and Mah (1982) propose that the endpoint of a continuum is naturally adopted as a cognitive reference point, and "can transform the underlying continuum, stretching the perceived distances between stimuli in the vicinity of a reference point relative to distances between stimuli far from it" (page 346). Thus, subjects apparently do not assign an equal weight to each possible colored bingo chip proportion, but instead put a greater weight on proportions near the (interior) reference point than on proportions farther from it, except when the salience of that point is broken in the three-color frame.

But why does the interior endpoint appear to weigh more heavily as a reference point than the exterior endpoint? We suspect the salience of the interior endpoint b is enhanced by a well-known proportionality bias in perception. Spence (1990) finds that subjects facing a partioned object (a number, line segment, box, or pie chart) typically perceive the smaller (larger) piece in greater (lesser) proportion than its objective value. Slusser and Barth (2017) report that subjects (children and adults) who are asked to physically locate a number's location on a line interval typically over- (under-) shoot its true value when that number is less than (greater than) the midpoint of the line segment. Thus, our minds tend to exaggerate the proportion reflected in smaller components. This bias forms a perceptual underpinning of the probability weighting function in cumulative prospect theory (Tversky and Kahneman, 1992), in which small (large) probabilities loom larger (smaller) than their objective values, as well as of a-sensitivity over binary outcomes derived from two states.

The remainder of the paper is organized as follows. Section II provides details of our experimental design. Section III develops hypotheses which can be directly addressed through implementation of our design. Section IV reports our findings. Section V summarizes our findings and describes potential applications.

II Experimental Design

Our design is implemented in three sequential phases and comprised of seven total treatments. All the sessions take place at Baruch College, City University of New York, except for Treatment 5, which takes place at Ridgewood Avenue School in Glen Ridge, New Jersey. Each participant in the experiment receives a $10 show-up payment, plus the possibility of additional cash rewards. We observe the incentivized choices of 439 subjects in Treatments 1-3 and 6. In addition, we observe the urn-filling choices of 151 subjects in Treatments 1 and 3-7, who receive a flat payment of $20 (including their show-up payment). Our design is summarized in Table 2.

In the first phase, Treatments 1-3, we attempt to elicit the perceived probabilities of DMs. Upon observing unexpectedly strong regularities in urn-filling choices in Treatments 1 and 3, we developed the second phase of our design, Treatments 4 and 5, in which all subjects fill Treatment 1 urns "for use in future experiments." In Treatment 4, the urn-filling decisions are made by college

Table 2: **Treatment summary**

Treatment	Mechanism	Ambiguity Source	Choice Subjects	Fill Subjects
1	BDM	Human Subject	122	15
2	BDM	Lottery	109	—
3	MPL	Human Subject	106	10
4	Chips-College	Human Subject	—	45
5	Chips-Kids	Human Subject	—	21
6	Binary	Human Subject	102	10
7	Chips-Binary	Human Subject	—	60

students, and in Treatment 5 they are made by fourth grade students. The focal treatment described in Section I, Treatment 6, is actually in the third phase of our design. Having concluded from Treatments 1-5 a strong, unexpected link between incentivized choices and urn-filling choices, Treatment 6 presents a simplified binary choice setting to robustly identify the pattern across the two Ellsberg (1961) elicitation frames, with a prior hypothesis about this link. In Treatment 7, also in the third phase of our design, we run related urn-filling-only sessions, in which subjects fill the Treatment 6 urns "for use in future experiments."

II.1 Treatment 1

In Treatment 1, subjects face a sequence of 20 decision problems. Each problem pays either \$0 or \$20. For the first 18 decision problems, a subject reports the minimum (certain) probability of winning \$20 which she prefers to a random draw from the given ambiguous bet. We interpret this reported probability as a revealed CEP, and elicit it using a variation on the BDM mechanism. The last two decision problems are the Ellsberg (1961) three-color problems, E.I and E.II. At the end of the experiment, one decision problem is selected at random, by draw from a bingo cage with numbered balls, to determine a subject's payment for the experiment.

These 20 decision problems are presented to the subjects in three sequential sets. Decision set L (for Large) is comprised of nine questions related to urns L1-L9, filled with a total of 100 bingo chips each. Decision set S (for Small) is comprised of nine questions related to urns S1-S9, filled with a total of 10 bingo chips each. Decision set E is comprised of the two Ellsberg three-color questions E.I and E.II. There are two bingo chip colors (blue and orange) in each of the urns L1-L9 and S1-S9, and three bingo chip colors (gray, blue, and orange) in urn E. Winning and losing colors are designated by subtreatment. A winning color is worth \$20 and a losing color is worth \$0. The feasible range of the number of winning chips in urns S1-S9 and L1-L9 is presented in Table 3. In urn E there are 3 gray chips and the remaining 6 chips are blue and orange in an unknown proportion.

Motivating our use of both small and large urns, Izhakian (2018) presents a theoretical model in which perceived ambiguity is greater when the discretization of a given probability range is less dense, implying a stronger effect on perceived probabilities in small urns relative to large urns. In turn, this implies that the deviation of perceived probabilities from the average probabilities in small urns is greater than in large urns. Filiz-Ozbay et al. (2018) report evidence that subjects prefer large over analogous small urns, a result driven by the subset of subjects who reveal themselves to be ambiguity averse. Pulford and Colman (2008) report evidence that revealed preferences are not affected by urn size.

Table 3: **Available urns**

Small Urns			Large Urns		
Urn	Winning Chips	Expected Winning Probability	Urn	Winning Chips	Expected Winning Probability
S1	0-2	0.1	L1	0-20	0.1
S2	0-4	0.2	L2	0-40	0.2
S3	0-6	0.3	L3	0-60	0.3
S4	0-8	0.4	L4	0-80	0.4
S5	0-10	0.5	L5	0-100	0.5
S6	2-10	0.6	L6	20-100	0.6
S7	4-10	0.7	L7	40-100	0.7
S8	6-10	0.8	L8	60-100	0.8
S9	8-10	0.9	L9	80-100	0.9

At the beginning of the experiment, one subject is selected at random (by draw from a bingo cage filled with numbered balls) to fill urns L1-L9, S1-S9, and E with colored bingo chips, in a separate room. The subject knows the urns will be used by others in the experiment, but is not informed that there are "winning" and "losing" colors, or that the other subjects will be betting on these urns. The rules for filling each urn are taped to the urn itself.[3] This subject is paid a

[3]For example, on L3, in subtreatments in which blue (unknown to the urn filler) is the winning color, the following text is attached: "You will place a total of 100 chips in this container. Between 0-60 chips must be blue, and the rest must be orange. One of the investigators will verify this container meets the color requirements (0-60 blue chips plus 40-100 orange chips in any combination so that the total number of chips is 100), but will offer no further instruction."

total of $20 (including show-up payment).

For each ambiguous urn in decision sets L and S, each of the remaining subjects is asked to report the minimum (certain) number of winning bingo chips in an alternative risk-only urn such that she prefers a random draw from the risk-only urn over a draw from the given ambiguous urn. That is, the subject reports a discretization of her CEP. Subjects are incentivized to reveal their CEPs through the BDM mechanism, as follows.[4] If the current decision problem is selected for payment, a random integer p between 0 and 100 is drawn from a cage of numbered ping pong balls. This number p represents the number of winning chips out of 100 in the alternative risk-only urn. If p is less than her reported (discrete) CEP, the subject draws from the ambiguous urn to determine her payment for the experiment; otherwise, she draws from the alternative risk-only urn with exactly p winning bingo chips out of 100 (which is constructed at that moment).

For urn E, with three gray chips and six blue and orange chips in an unknown proportion, subjects are asked two questions: (E.I) On a random draw from E, would you prefer the winning color to be gray or blue; and (E.II) On a random draw from E with two winning colors, would you prefer the winning colors be gray and orange, or blue and orange.[5]

There are four subtreatments within Treatment 1, labeled LSE-B, SLE-B, LSE-O, and SLE-O. The first three letters stand for the order of choice sets, and the final letter denotes the winning color (<u>B</u>lue or <u>O</u>range). For example, in LSE-B, subjects first face the Large (100-chip) urns, then the Small (10-chip) urns, and finally the Ellsberg urn, and blue is the winning color.[6] The order that urns are presented to subjects within each set L and S is always 3, 6, 8, 1, 5, 9, 2, 7, 4; e.g., the urn with a feasible range of 0-20% winning chips is presented to subjects fourth.[7] Each subtreatment has an approximately equal number of subjects.

[4]BDM mechanism is most commonly used to elicit certainty equivalent outcome values for risk-only bets, although Halevy (2007) elicits certainty equivalent outcome values for ambiguous bets. A variant of BDM for probabilities is applied in Grether (1981) and Crockett and Crockett (2019). Karni (2009) formalizes subjective probabilities elicitation over ambiguous payoff distributions using this mechanism.

[5]As written, these questions coincide with subtreatments where blue is the winning color in L1-L9 and S1-S9. In the other subtreatments, E.I is a choice between gray or orange, and E.II is a choice between gray and blue, or orange and blue.

[6]In decision set E, in the B subtreatments, the first choice is between gray and blue, and the second is between gray and orange, or blue and orange. In the O subtreatments, the first choice is between gray and orange, and the second is between gray and blue, or orange and blue.

[7]We use scrambled order following a pilot session which reveals a sharp drop in the mean/median CEP between urns 5 and 6 (both sizes) when urns are presented in monotonic order (1, 2, 3, 4, 5, 6, 7, 8, 9). Ex post, the distribution of CEPs in Treatment 1 compared to the pilot is indistinguishable except in L5 and S5, where the CEPs are greater in the pilot than in Treatment 1.

II.2 Treatment 2

Treatment 2 is identical to Treatment 1, with the same subtreatments, except that the source of ambiguity is not generated by a human subject. Rather, the number of blue and orange bingo chips in urns L1-L9, S1-S9, and E is generated by a uniform random device. Specifically, the number of winning chips for a given urn is drawn from a cage of numbered ping pong balls, restricted to the allowable range, *after* subjects report their CEPs for each urn. Thus, "ambiguity" in this case is practically a compound lottery, and subjects are explicitly informed of this procedure by which urns will be filled.

There are many options to implement an alternative source of ambiguity to the one used in Treatment 1. Critically for our purposes, we do not want a source whose data generating process (DGP) may objectively match a non-uniform DGP that the urn-filling subjects in Treatment 1 may implement. If the CEPs in Treatment 2 are similar to those in Treatment 1, while the underlying DGP is not, it is unlikely that the objective uniform DGP itself is influencing second-order beliefs about the underlying probability distributions.

II.3 Treatment 3

The use of BDM mechanism to elicit CEPs raises the concern that some subjects may face difficulty understanding the mechanism (Plott and Zeiler, 2005). To test the robustness of our findings to an alternative elicitation mechanism, in Treatment 3, we use the MPL mechanism to elicit CEP, as in Charness et al. (2013).[8] This mechanism can be interpreted as "a discrete implementation of BDM through a sequence of pairwise choices" (Freeman et al., 2018), and is generally considered less cognitively demanding of subjects than BDM (Harrsion and Rutström, 2008).

For a given decision problem, MPL is presented as a list of binary choices over a range of alternative probabilities, with one choice selected at random to implement for payment. For example, consider L3, in which the chance of winning \$20 is between 0 and 60%. In BDM, the CEP is elicited directly on a grid of 1 percentage point increments between 0 and 60. The corresponding MPL is implemented on 2 percentage point grid with fifteen gridpoints, centered on the midpoint of the probability range. The first of fifteen binary choices is between a draw from the ambiguous urn (L3) and a draw from a risk-only urn with an exactly 16% chance of winning. The second choice

[8] A third technique commonly used in psychology for eliciting indifference, *titration*, is implemented by Binmore et al. (2012)

is between a draw from L3 and a draw from a risk-only urn with an 18% chance of winning. And so on, through the fifteenth choice, where the choice is between a draw from L3 and a draw from a risk-only urn with a 44% chance of winning. This creates a 28 percentage-point range of risk-only alternatives that are symmetrically and uniformly distributed about 30%—the midpoint of the winning probability range of L3. If a subject chooses L3 at low probabilities and then switches to the risk-only alternative for all probabilities above some subjective threshold probability, we interpret that threshold as a discrete approximation of the subject's CEP. In section IV, we address subjects who switch multiple times.

Both BDM and MPL have their relative advantages. MPL may be easier for subjects to understand. However, with BDM the CEP of a bet can be elicited by a single choice, while in MPL the same bet requires fifteen binary choices to elicit the CEP on a grid twice as course.

Treatment 3 focuses on urns L3, L5, and L7, defined in Treatment 1. Sprenger (2015) reports that, at least in the domain of risk, the fixed lottery in MPL serves as an endowment, and revealed preferences are strongly biased in favor of that endowment. Therefore, to address concern of a possible endowment effect for ambiguity, we also implement an "endowment reversal" MPL for each of these urns. Instead of fixing the ambiguous urn and eliciting a switching point for a range of risk-only urns with increasing probabilities, we fix a risk-only urn and provide fifteen alternative ambiguous urns with an increasing range of winning probabilities. The risk-only urn L3R has 30 winning chips out of 100; L5R has 50 winning chips out of 100; and L7R has 70 winning chips out of 100. The winning range in the alternative ambiguous urn increases monotonically by 4 percentage points from alternative urn 1 to alternative urn 15.[9] The eighth (middle) ambiguous alternative is identical to L3 for risk-only urn L3R, L5 for L5R, and L7 for L7R. For each of the six CEP elicitations (L3, L5, L7, L3R, L5R, and L7R), the fifteen alternative urns are presented simultaneously as a list of fifteen binary choices sorted by probability range.

In addition to the six MPLs, subjects in Treatment 3 also face three binary choice extensions of the three-color Ellsberg question E.I. In decision problem E1.I, urn E1 is filled according to the same rules as urn E in Treatments 1 and 2, but subjects decide which of the three colors is the winning color rather than being restricted to two colors as in E.I. This specification eliminates concern that subjects believe the available ambiguous color choice in E.I is adopted strategically

[9]For example, for risk-only urn L5R, the alternative urns, in sequence, have winning ranges of 0-72, 0-76, 0-80, 0-84, 0-88, 0-92, 0-96, 0-100, 4-100, 8-100, 12-100, 16-100, 20-100, 24-100, and 28-100.

by the experimenter; although Oechssler and Roomets (2015) report that subjects do not appear to distinguish between purely mechanical ambiguity and experimenter-generated ambiguity. In decision problem E2.I, there are two urns: urn E2-1 is governed by the same color rules as urns E and E1, while urn E2-2 contains 9 chips, 0-6 of which are blue and the rest are gray or orange in an unknown proportion. Gray is the winning color in urn E2-1, while blue is the winning color in urn E2-2. Each subject is asked to choose from which urn she would prefer a random draw. Thus, in E2.I, the subject chooses between gray (winning probability $\frac{1}{3}$) and blue (winning probability between 0 and $\frac{2}{3}$), just as in E.I, but these available choices are split between two urns rather than contained within one. Finally, in decision problem D.I (the two-color two-urn dual to E.I), subjects choose between a draw from urn 1, with 0-6 winning chips out of 9 chips, or from urn 2, with exactly 3 winning chips out of 9 chips (both of these urns have only two colors).

In summary, in Treatment 3 subjects face six (15-choice) MPLs and three Ellsberg-related binary choices, for a total of 93 binary choices. At the end of the experiment, for each subject, one of these 93 choices is selected at random by a draw from a cage of numbered ping pong balls to determine which bet she will actually play. In all MPLs, blue is the winning color. We use four subtreatments to vary the order of the choice sets in this treatment: MPL-1, MPL-2, MPL-3, and MPL-4. In each subtreatment, subjects begin with an Ellsberg question; then move to three MPLs (either L5 then L3 then L7, or L5R then L3R then L7R); then move to another Ellsberg question; then move to the remaining three MPLs (in the same order as stated above); and finally finish with the remaining Ellsberg question. The specific order in each subtreatment is presented in Table 4.

Table 4: **Decision sets by subtreatment in Treatment 3**

Choice Set	MPL-1	MPL-2	MPL-3	MPL-4
1	E2.I	D.I	E1.I	E2.I
2	L5	L5	L5R	L5R
3	L3	L3	L3R	L3R
4	L7	L7	L7R	L7R
5	D.I	E1.I	E2.I	D.I
6	L5R	L5R	L5	L5
7	L3R	L3R	L3	L3
8	L7R	L7R	L7	L7
9	E1.I	E2.I	D.I	E1.I

The source of ambiguity for all decision sets is a subject selected at random at the beginning of the session, as detailed in Treatment 1. Urns L3, L5, L7, L3R, L5R, L7R, and the Ellsberg urns are all physically constructed. The 45 alternative ambiguous urns corresponding to the risk-only endowment urns are constructed virtually. For each of these urns, we ask the urn-filling subject to record how many blue and orange chips she would place in the urn. When one of these alternative ambiguous urns becomes necessary to determine a subject's payment, the urn is physically constructed according the record of the urn filler. Subjects making incentivized choices are informed that these ambiguous alternative urns are not actually being filled by the urn filler, but are rather filled as needed according to the urn filler's record.

II.4 Treatment 4

In Treatment 4, to assess the robustness of urn-filling choices in Treatments 1-3, we elicit only urn-filling decisions. To this end, we replicate the urn-filling task of the randomly selected subject in Treatment 1. Subjects in Treatment 4 are invited to fill the urns in L, S, and E with colored bingo chips according to the same rules as in Treatment 1, but for "an upcoming experiment," rather than for other subjects recruited in the same session. We implement this treatment in order to collect sufficient urn-filling data for statistical hypothesis tests.

II.5 Treatment 5

Treatment 5 is a replication of Treatment 4, but the subjects are students in a *fourth grade* classroom at Ridgewood Avenue School in Glen Ridge, New Jersey, rather than Baruch College students. We chose fourth grade students in consultation with the principal and teachers at the school, who view this grade as comprised of the youngest students who can handle the task with minimal errors. Rather than pay each subject $20 individually, we opted instead for a group reward: the classroom teacher was presented with a check for $500 to spend on fun items/activities for the classroom. The size of the class is 25 students, but only 21 were present for the experiment.

II.6 Treatment 6

In Treatment 6, each subject makes 13 binary choices: E.I, E.II, D.I, and D.II, along with binary choice versions of S1-S9. For S1-S9, urns 1 and 2 are comprised as specified in Table 3, and subjects simply choose from which urn to draw. As in Treatments 1 and 2, the order of the questions is

S3, S6, S8, S1, S5, S9, S2, S7, S4. Questions S1-S9 are sandwiched between the Ellsberg three-color questions and their duals. In subtreatment ESD, the experiment begins with the three-color Ellsberg choices and ends with their two-urn duals, and in DSE the order is reversed. Thereby, in Treatment 6 we can analyze the impact of the elicitation frame on the Ellsberg questions within-subject, and also restrict focus to the first two questions and compare the framing impact across subjects (in ESD vs. DSE) without concern for order effects. Since we find no evidence of color effects in earlier subtreatments, blue is the winning color in both subtreatments. The urn filler is selected randomly and managed similarly to Treatments 1 and 2.

II.7 Treatment 7

Treatment 7 replicates the urn-filling tasks of the randomly selected subject in Treatment 6. Subjects in Treatment 7 are requested to fill the urns with colored bingo chips according to the same rules as in Treatment 6, but for "an upcoming experiment," rather than for other subjects recruited in the same session.

III Hypotheses

We develop several testable hypotheses from our experimental design related to revealed preferences and urn-filling choices. As discussed in Section I, revealed preferences might be subject to distinctions between likely and unlikely gains. To define likely and unlikely gains, consider a binary outcome bet in which the "winning" (larger) outcome (payoff) has an uncertain probability in the range $[\ell, u] \subseteq [0, 1]$. Let $m = \frac{1}{2}(\ell + u)$ be the uniform expected probability of winning; that is, m is the midpoint of the range of winning probabilities. When $m < \frac{1}{2}$ we refer to the bet as an *unlikely gain*, and when $m > \frac{1}{2}$ we refer to the bet as a *likely gain*.

As reviewed by Trautmann and van de Kuilen (2015), there is a growing consensus in the experimental literature (e.g., Curley and Yates, 1985, 1989; Di Mauro and Maffioletti, 2004; Armantier and Treich, 2015; Dimmock et al., 2016; Kocher et al., 2018) and the empirical literature (e.g., Brenner and Izhakian, 2018) about the different attitudes toward ambiguity concerning likely and unlikely gains. Our first two hypotheses follow suit.

Hypothesis 1 ***Ambiguity seeking for unlikely gains:*** *Given an ambiguous bet whose winning probability range has midpoint* $m < \frac{1}{2}$*, the median subject prefers the ambiguous bet to a risk-only bet with winning probability equal to* m*.*

Hypothesis 2 ***Ambiguity aversion for likely gains:*** *Given an ambiguous bet whose winning probability range has midpoint* $m > \frac{1}{2}$*, the median subject prefers a risk-only lottery with winning probability equal to* m *to the ambiguous bet.*

The next hypothesis considers the impact of the source of ambiguity on individuals' perceived probabilities (CEP) and thereby on their incentivized choices. That is, whether the ambiguous probability being chosen by a human being or by a uniform randomization device affects CEPs. Based upon existing literature (e.g., Halevy, 2007; Oechssler and Roomets, 2015; Armantier and Treich, 2015), we do not expect the source of ambiguity to substantially affect revealed preferences.

Hypothesis 3 ***Source independence ambiguity:*** *CEPs are independent of the source of ambiguity.*

The next hypothesis considers the impact of urn size on CEPs.

Hypothesis 4 ***Size effect:*** *CEPs will deviate further from the midpoint,* m*, of the winning probability range in the small urns relative to their large urn counterparts.*

We also expect to observe an endowment effect in Treatment 3.

Hypothesis 5 ***Endowment effect:*** *When facing a binary choice between an ambiguous bet and a risk-only bet within MPL, more subjects prefer the ambiguous bet to the risk-only bet when the ambiguous bet is framed as the endowment than when the risk-only bet is framed as the endowment.*

We turn now to the choices made by urn fillers about the number of bingo chips of each color to place in the urns. Our initial hypothesis, prior to collecting any data, is that subjects are not systematically biased in their fill choices.

Hypothesis 6 ***Urn-filling choices:*** *Urn-filling choices are symmetrically distributed about the midpoint,* m*, of the winning probability range.*

Contrary to Hypothesis 1, Ellsberg (1961) conjectures that, in choice E.I of the three-color experiment, subjects act as if they are ambiguity averse for unlikely gains. While the early experimental literature supports this conjecture (surveyed in Camerer and Weber, 1992), more recent studies find the revealed preference for ambiguity in E.I to be weak or neutral (e.g., Binmore et al., 2012; Charness et al., 2013; Stahl, 2014). We hypothesize that in the two-urn frame of D.I, revealed preferences for ambiguity are stronger relative to E.I, producing behavior consistent with Hypothesis 1. In Treatment 3, choices E1.I and E2.I are designed as intermediate decompositions from E.I to D.I. Therefore, we expect revealed preferences favoring ambiguity to increase incrementally from E.I to D.I, so that the fraction of subjects who prefer the ambiguous urn to the risk-only urn satisfies E.I < E.1.I < E.2.I < D.I. Note that these Ellsberg urns share the same range of winning probabilities for unlikely gains.

Only after analyzing the results from Treatments 1-5 did we develop decision problem D.II, the dual to E.II. Given the observed impact of both states and consequences on choices in E.I and D.I in Phase 1, we expect to observe behavior consistent with ambiguity aversion for likely gains in D.II, but to a significantly greater extent than for E.II. Thus, we hypothesize the rank of the Ellsberg-type choices with respect to favoring risk-only bets will be as follows.

Hypothesis 7 ***Ellsberg ordering:*** *The fraction of subjects who prefer ambiguity to risk-only in the Ellsberg choices aligns in the following order, from smallest to greatest: D.II, E.II, E.I, E1.I, E2.I, D.I.*

Further, in light of results from Phases 1 and 2, in Phase 3, we expect that urn-filling subjects in D.II, who are asked to fill a 9-chip urn with 3-9 blue bingo chips and the rest orange, will fill the urn with fewer than 6 blue chips (the midpoint of the winning range), a choice consistent with urn-filling choices for two-urn problems in Treatments 1, 3, 4, and 5.

Finally, we do not expect to observe color or order effects within subtreatments.

Hypothesis 8 ***No subtreatment effect:*** *Reversing the role of colors and the order of questions within a treatment do not affect choices.*

IV Findings

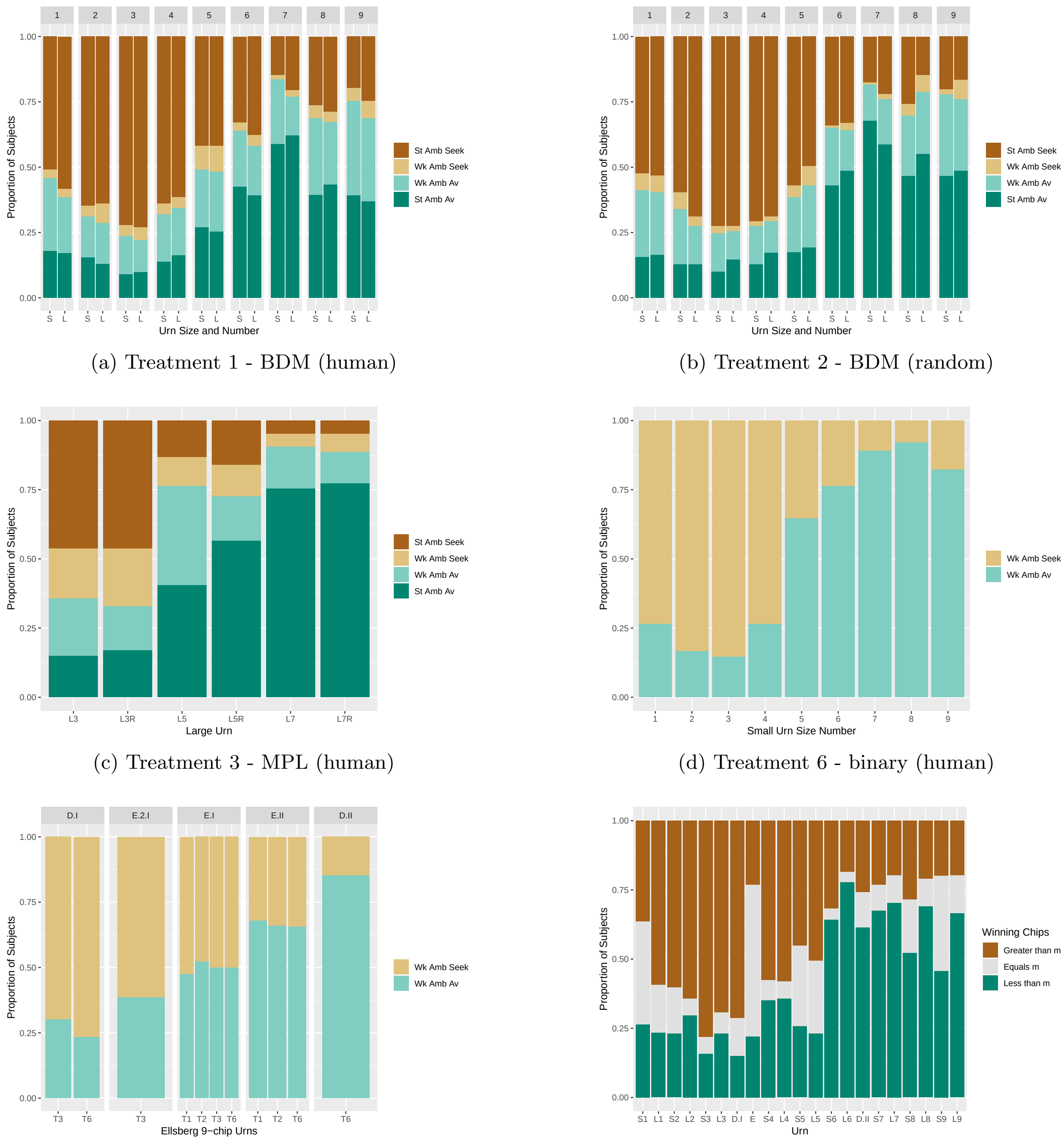
IV.1 Overview

We begin with an overview of our incentivized choice findings from Treatments 1, 2, 3, and 6. For each decision problem, we classify a subject's decision as consistent with one of four revealed preference types: strict ambiguity seeking, weak ambiguity seeking, strict ambiguity aversion, and weak ambiguity aversion. Figure 2 presents a visualization of this identification. In each subfigure, ambiguous urns are arranged in increasing order of the midpoint, m, of the range of winning probabilities.

We infer revealed preferences from the implied CEP in BDM (Figures 2a and 2b) and MPL (Figure 2c) decision problems. We expect a greater proportion of weak preferences in MPL than in BDM, given the increased discreteness of its elicitation grid. In BDM, a subject reports the minimum certain probability p of winning (on a one-percentage-point grid) for which she prefers the risk-only urn to the ambiguous urn with a midpoint m of winning probabilities. Thus, $p < m$ implies strict ambiguity aversion (underweighting the probability of the favorable outcome); $p = m$ implies weak ambiguity aversion; $p = m + .01$ implies weak ambiguity seeking; and $p > m + .01$ implies strict ambiguity seeking (overweighting the probability of the favorable outcome).

In MPL, if a subject switches her choice between the ambiguous urn and its risk-only alternative at most once as the winning probability of the latter increases, we can infer the switching probability as her CEP. This is the case for most of our subjects.[10] For a subject who switches multiple times, we cannot infer a CEP directly; we adjust the CEP similarly to how non-monotonic switching is handled in the risk preference elicitation literature (e.g., Holt et al., 2002). First, we count the number of risk-only bets (of fifteen possible) chosen by the subject in a given list, denoted r. Since our MPL is implemented on a 2-percentage-point grid, each unit deviation from $r = 8$ (which corresponds to CEP $= m$) changes our adjusted CEP measure by two percentage points. Thus, our adjusted CEP measure is defined as $m + .02 \times (8 - r)$. The same measure is also utilized to identify subject preference types as specified above for BDM, where .01 is replaced with .02 in the

[10] On average, about 18% of the subjects switch multiple times in decision problems L3, L5, and L7, and 29% switch multiple times in decision problems L3R, L5R, and L7R. The proportion of multiple switchers in L3, L5, and L7 are consistent with behavior observed in other MPL experiments that elicit preferences for ambiguity (e.g., Charness et al., 2013) and preferences for risk (e.g., Holt et al., 2002). The proportion of multiple switchers in L3R, L5R, and L7R is a bit high, perhaps reflecting increased complexity of this task.



(a) Treatment 1 - BDM (human)

(b) Treatment 2 - BDM (random)

(c) Treatment 3 - MPL (human)

(d) Treatment 6 - binary (human)

(e) Ellsberg questions

(f) Urn fillers

Figure 2: **Revealed preferences**

definitions of weak and strict attitude toward ambiguity.[11]

Each Ellsberg decision problem E.I, E.1.I, E2.I, E.II, D.I, and D.II, along with S1-S9 in Treatment 6, offers a binary choice between an ambiguous urn and a risk-only alternative with a winning

[11] Assigning fractional types, which sum to one for subjects who switch non-monotonically, generates a figure virtually indistinguishable from Figure 2c.

probability equal to m. These choices allow only for characterizing weak preferences, which are presented in Figures 2d and 2e.[12] Figures 2a-2e suggest that, indeed, subjects do tend to seek ambiguity for unlikely gains and are averse to ambiguity for likely gains. This holds true across all urns and elicitation mechanisms, *except* in Ellsberg's three-color choice 1 (E.I in Figure 2e), where subjects are equally split between choosing risk-only urn and ambiguous urn across all treatments.

Finally, before turning to formal statistical tests of our hypotheses, we note that Figure 2f presents a visualization of urn-filling choices partitioned into three groups: fill choices above, below, and equal to the midpoint of the range of possible winning chips. These findings are discussed in detail following the analysis of revealed preferences for incentivized DMs. Notice, however, that in this figure urn E appears to be the exception to the rule governing other urns, breaking the tendency of choices to be located between m and the interior boundary of the range of winning chips.

IV.2 Revealed preferences for ambiguity concerning likely and unlikely gains

Table 5 reports the mean and median revealed CEPs using the BDM and MPL mechanisms, along with the proportion of subjects who prefer the ambiguous urn to the risk-only urn in binary decision problems (Treatment 6 for S1-S9, and pooled choices, across all relevant treatments, for E.I, E.II, D.I, and D.II). It also presents urn-filling choices, pooled across treatments, expressed as the fraction of winning chips placed in the urns, as discussed in Subsection IV.4.

It can be observed from Table 5 that for each urn with unlikely gains, the mean and the median CEP in Treatments 1, 2, and 3 are greater than the midpoint m, and for each urn with likely gains, the mean and the median CEP are less than m. Further, in Treatment 6, in S1-S9 and the Ellsberg decision problems E.II, D.I, and D.II, when facing unlikely gains the median subject prefers the ambiguous urn, and when facing likely gains the median subject prefers the risk-only urn. Thus, the findings are uniformly consistent with Hypotheses 1 and 2, except for E.I.

To test Hypotheses 1 and 2 statistically in their weak and strict forms, we utilize the binomial test. This test is conservative, requiring only independence of choices between subjects, and can be applied to choices made within each mechanism. To test Hypothesis 1 (ambiguity seeking for unlikely gains) for strict preferences for a BDM or MPL decision problem with grid size g (.01 and

[12] Figures 2d and 2e pool the choices in E.I and E.1.I, since allowing subjects to choose their preferred winning color in E.1.I does not impact the distribution of choices relative to E.I.

Table 5: **Small and large urns**

Urn	m	Tr.1 (Tr.2) Mean CEP	Tr.1 (Tr.2) Med CEP	Pooled Tr.1-2 Mean CEP	Pooled Tr.1-2 Med CEP	Tr.3 Mean CEP	Tr.3 Med CEP	Tr.6 (Pooled) Choose Ambg	Pooled Mean Fill	Pooled Med Fill
S1	.1	.127 (.124)	.12 (.12)	.126	.12	NA	NA	.74	.110	.10
L1	.1	.129 (.124)	.13 (.12)	.127	.12	NA	NA	NA	.129	.15
S2	.2	.263 (.242)	.25 (.25)	.254	.25	NA	NA	.83	.262	.30
L2	.2	.253 (.257)	.25 (.25)	.255	.25	NA	NA	NA	.261	.30
S3	.3	.394 (.386)	.40 (.40)	.391	.40	NA	NA	.85	.442	.50
L3	.3	.399 (.396)	.40 (.40)	.398	.40	.344	.32	NA	.392	.43
S4	.4	.490 (.487)	.50 (.50)	.489	.50	NA	NA	.74	.448	.50
L4	.4	.486 (.480)	.46 (.48)	.483	.48	NA	NA	NA	.466	.50
S5	.5	.526 (.549)	.51 (.55)	.537	.55	NA	NA	.35	.593	.50
L5	.5	.535 (.553)	.51 (.51)	.543	.51	.489	.50	NA	.628	.51
S6	.6	.576 (.573)	.60 (.60)	.575	.60	NA	NA	.24	.530	.50
L6	.6	.598 (.591)	.60 (.60)	.594	.60	NA	NA	NA	.462	.40
S7	.7	.642 (.637)	.65 (.65)	.639	.65	NA	NA	.10	.609	.50
L7	.7	.650 (.651)	.65 (.65)	.651	.65	.637	.62	NA	.593	.51
S8	.8	.784 (.775)	.80 (.80)	.780	.80	NA	NA	.08	.760	.70
L8	.8	.783 (.759)	.80 (.76)	.771	.80	NA	NA	NA	.721	.68
S9	.9	.889 (.877)	.90 (.90)	.883	.90	NA	NA	.18	.874	.90
L9	.9	.894 (.878)	.90 (.90)	.886	.90	NA	NA	NA	.861	.83
D.I	$\frac{1}{3}$	NA	NA	NA	NA	NA	NA	(.73)	.47	.56
E.I	$\frac{1}{3}$	NA	NA	NA	NA	NA	NA	(.50)	.35	.33
E.II	$\frac{1}{3}$	NA	NA	NA	NA	NA	NA	(.33)	.35	.33
D.II	$\frac{1}{3}$	NA	NA	NA	NA	NA	NA	(.15)	.57	.56

.02, respectively), the null hypothesis is that for unlikely gains the median subject reports a CEP less than or equal to $m + g$. Therefore, we reject the null hypothesis if the proportion of subjects with a CEP greater than $m + g$ is significantly greater than .5. To test Hypothesis 1 for weak preferences, our null hypothesis is that the median CEP is less than or equal to m. Similarly, to test Hypothesis 2 (ambiguity aversion for likely gains) in strict (weak) form, our null hypothesis is that for likely gains the median CEP is greater than or equal to $m - g$ (m).

As there is a strong similarity between the two BDM treatments, apparent in Figures 2a and 2b along with Table 5, our BDM tests focus on data pooled between Treatments 1 and 2. Pooling these treatments is justified statistically in Subsection IV.3. For the large urns, the strict ambiguity-seeking preference in Hypothesis 1 is supported well below the 1% confidence level for L2-L4, and at the 5% confidence level for L1. The strict ambiguity-aversion preference in Hypothesis 2 is supported well below the 1% confidence level for L7, and the weak preference is supported for L6-L9 below the 1% confidence level. The same conclusions hold true for the small urns, except the weak ambiguity-seeking preference in Hypothesis 1 is only confirmed for S1 at the 5% confidence level.

Within MPL, only the weak ambiguity-seeking preference in Hypotheses 1 is confirmed, well below the 1% confidence level for both endowment conditions (L3 and L3R). The strict ambiguity-aversion preference in Hypothesis 2 is also confirmed for both endowment conditions (L7 and L7R). For binary choices S1-S9 (Treatment 6), where only weak preferences can be tested, we confirm Hypotheses 1 and 2 for all urns far below the 1% confidence level.

For the Ellsberg questions, we confirm Hypotheses 1 and 2 well below the 1% confidence level for D.I, E.II, and D.II in all treatments, individually and pooled. We confirm Hypothesis 1 for E.2.I at the 2% confidence level. Responses to decision problems E.1.I and E.I are not significantly different from a fair coin flip within any treatment individually, nor when pooled across treatments in the case of E.I (E.1.I is only presented in Treatment 3). Recall that E.1.I is identical to E.I, except that in E.1.I the subject may choose any one of the three colors to be the winning color, rather than being restricted to the risk-only color and one of the two ambiguous colors as in E.I. In both E.I and E1.I, across treatments, one-half of the subjects choose the risk-only color: 53 of 106 subjects choose the risk-only color in E.1.I, and 166 of 333 choose the risk-only color in E.I. Therefore, it does not appear that subjects believe the experimenter is biased in selecting the winning colors, or that the winning color is strategically picked, corroborating the findings in Oechssler and Roomets (2015).

Overall, our findings demonstrate very strong evidence for ambiguity seeking for unlikely gains and ambiguity aversion for likely gains across the three different elicitation mechanisms, except for decision problem E.I and its close substitute E.1.I, which we revisit after characterizing the urn-filling choices. These findings are in line with a-sensitivity (Dimmock et al., 2016) and the growing evidence from the experimental literature (summarized in Trautmann and van de Kuilen, 2015).

While we do not state prior hypotheses regarding response to fully ambiguous urns L5 and S5, we note that choices in these urns correspond to weak ambiguity aversion or seeking, depending on the mechanism. Subjects tend to be mildly averse to ambiguity for these urns in the binary choice and MPL formats, and mildly seeking for ambiguity in BDM. This matches an overall increase in ambiguity seeking in BDM mechanism relative to MPL. In BDM, this finding is not statistically significant. In MPL (L5 and L5R) and the binary choice version of S5, the finding is significant below the 1% confidence level under the two-tailed binomial test.

Finding 1 ***Ambiguity seeking for unlikely gains:*** *Consistent with Hypothesis 1, subjects reveal preferences consistent with ambiguity seeking for unlikely gains across all elicitation mechanisms.*

Finding 2 ***Ambiguity aversion for likely gains:*** *Consistent with Hypothesis 2, subjects reveal preferences consistent with ambiguity aversion for likely gains across all elicitation mechanisms.*

IV.3 Source-independent preferences for ambiguity

An important question in interpreting our findings is whether the source of ambiguity affects individuals' revealed preferences. Based on the existing literature, in Hypothesis 3, we conjecture that individuals are not affected by the source of ambiguity. In Treatment 1, the source of ambiguity is a human being, while in Treatment 2 the source is a randomization device, otherwise these two treatments share the same design (BDM mechanism). Therefore, Hypothesis 3 can be tested by the significance of the statistical difference between the elicited CEPs in Treatments 1 and 2. We compare this difference, urn by urn, using the Mann-Whitney rank sum test, where the null hypothesis is that the two groups of CEPs are drawn from the same underlying distribution.

For the small urns, no distribution of CEPs is significantly different between Treatments 1 and 2 at the 5% confidence level, and only in S5 is the difference significant at the 10% confidence level. For the large urns, the distributions of CEPs in L8 and L9 are significantly different between Treatments 1 and 2 at the 5% confidence level (p-value of .023 for both L8 and L9), but the differences in distributions of the remaining seven urns are highly insignificant (all p-values for L1-L7 are in excess of .25). Thus, out of 18 CEP pairs, 15 are not significantly different at the 10% confidence level, and 13 pairs are not significantly different even at the 25% confidence level. Of the three significantly different CEP pairs (p-values between .02 and .08), it is the random source of "ambiguity" (Treatment 2) that causes the larger deviation of CEPs from midpoint m toward the interior boundary b of the range of winning probabilities. In no urn does the human being source of ambiguity induce a significantly stronger bias in CEPs than the post-decision uniform randomization. Thus, it is quite unlikely that the bias in revealed preferences relative to m, reported in Findings 1 and 2, is caused by subjects attempting to predict a bias in the urn-filling choices.

Finding 3 ***Source-independent ambiguity:*** *There is no supportive evidence that the systematic deviations from ambiguity-neutral choices are caused by attempts to predict the choices of urn fillers.*

IV.4 Urn-filling choices

To characterize urn-filling choices, we first establish the appropriateness of pooling our urn-filling data across treatments. In Treatments 1, 3, and 6, respectively, 15, 10, and 10 subjects fill urns for use by subjects in their own sessions. In addition, in Treatments 4 and 5, respectively, 45 college students and 21 fourth grade students fill BDM treatment urns for "future experiments" and, in Treatment 6, another 60 college students fill binary choice urns for "future experiments."

For each urn type, we test whether there is a significant difference between the distributions of urn-filling choices across treatments using the Kruskal-Wallis H test. We find no significant difference between the distributions of urn-filling choices for the small urns S1-S9 (Treatments 1, 4, 5, 6, and 7), even at the 15% confidence level. For the large urns, out of the nine choices, two (L7 and L9) have significantly different distributions by treatment (p-values .011 and .005, respectively). In Treatment 4, the college students filling urns L7 and L9 for future experiments tend to place more winning chips than subjects in other treatments. Nevertheless, there is no systematic over-placing bias by treatment (or over-placing for urns with $m > .5$).

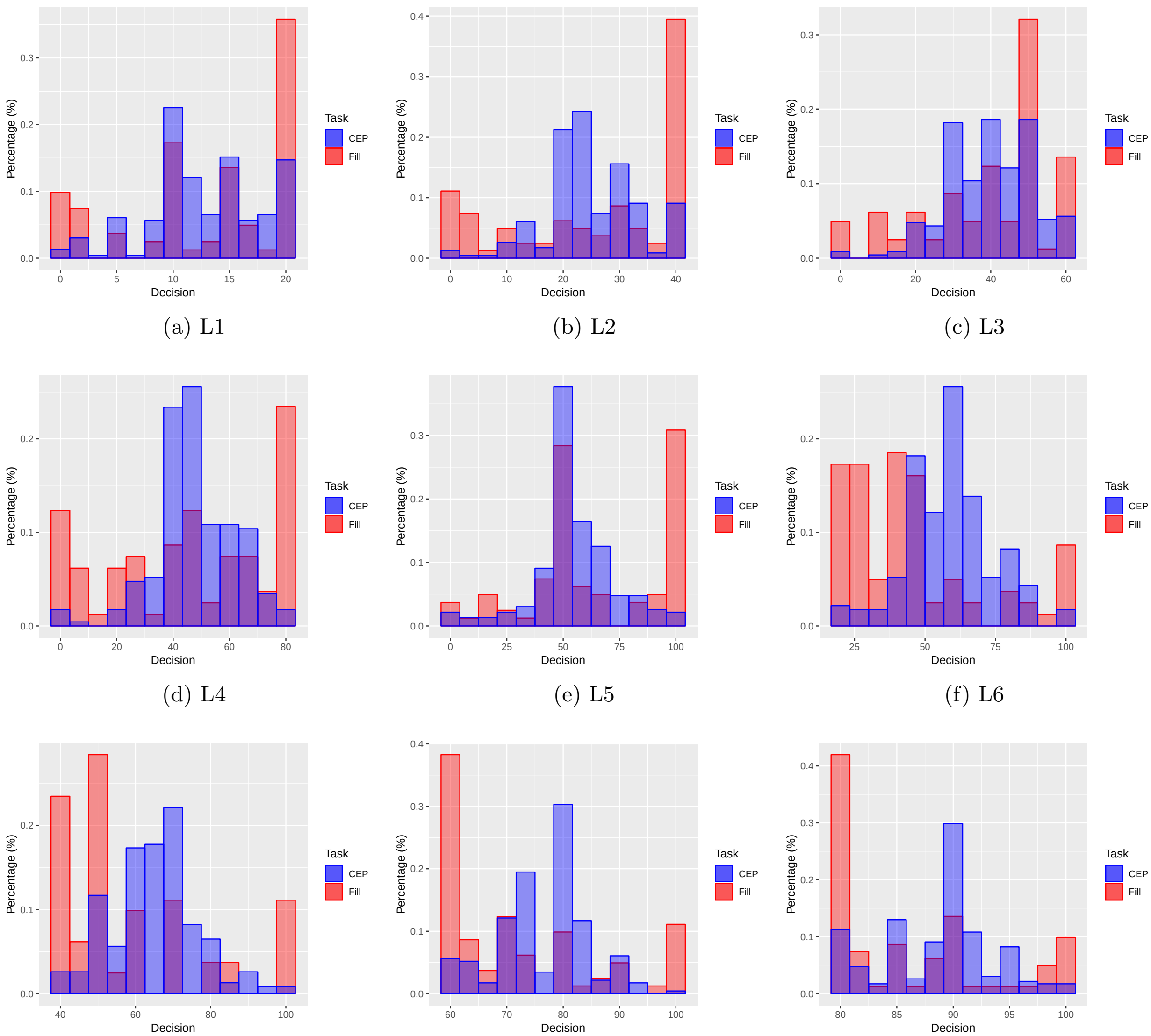
In the Ellsberg questions, there is no significant difference in the distributions of urn-filling choices for D.I and D.II across Treatments 3, 6, and 7. Across Treatments 1, 3, 4, 5, 6, and 7, there is a significant difference in the distributions of urn-filling choices for E. In Treatment 3, the ten subjects filling urn E.1.I, which is filled according to identical rules as E.I and E.II, exhibit more extreme fill choices than subjects in the other treatments. Nevertheless, including these subjects in the pooled sample runs counter to our main urn-filling findings that choices in E are tightly symmetric about m. Therefore, we use a pooled sample for our urn-filling analysis.

To facilitate comparison with the incentivized choices, the right two columns of Table 5 report the mean and median urn-filling choices, expressed as fractions of total chips in the urn. Contrary to Hypothesis 6, urn fillers appear to exhibit strong preferences. Aside from E, for each urn with unlikely gains, the mean urn-filling choice is greater than the midpoint of the winning chips range, m, and for each urn with likely gains the mean choice is less than m, sharing a similar pattern with incentivized choices. This shared pattern is especially evident in Figure 2, as the graph representing urn-filling choices, 2f, is remarkably similar to its incentivized choices counterparts.

In a given urn, the median urn-filling choice tends to be further from the midpoint m than the mean, suggesting that these distributions are skewed with long tails extended toward the exterior boundary of the winning probability range, as illustrated in Figure 4. The histograms for urn-filling choices in L1-L9, presented in Figure 3, confirm this. For the sake of comparison, this figure also includes the corresponding CEPs. There is a clear mode for urn-filling choices on the interior boundary of the winning probability range, and when an equal number of both colors is feasible (in L3-L7), these choices present an additional mode equal to the midpoint of the feasible probability range, m. Thus, while the distribution of CEPs for every urn appears to be symmetric (with mean between m and b, closer to m), the distributions of urn-filling choices are clearly not.

Under the one-sample sign test, all large urns (pooled across Treatments 1, 3, 4, and 5) have median fill choices that are significantly different from the midpoint of the winning probability range, m, at the 1% confidence level, except for L4, which has a p-value of .0505. For the large urns, the median fill choice is significantly greater than m for L1-L5, and significantly less than m for L6-L9. For the small urns (pooled across Treatments 1, 4, 5, and 6), all median fill choices are significantly different than m at the 1% confidence level, except for S1 whose median choice insignificantly differs from m (p-value .151). Similarly to the large urns, the median choice for the small urns is greater than or equal to m for S1-S5, and less than or equal to m for S6-S9. Thus, the distribution of choices in both small and large urns has a central tendency in the direction of the interior boundary of the range of probabilities, b, relative to m.

All Pearson skewness coefficients are negative for urns S1-S4 and L1-L4, averaging about -.6 in both cases, and Pearson skewness coefficients are positive for urns S6-S9 and L6-L9. We adopt the following significance test of skewness. For a given urn with an even number of choices n, partition the choices into two groups: group H consists of the highest $\frac{n}{2}$ choices and group L consists of the lowest $\frac{n}{2}$ choices, where choices are expressed as a fraction of the total chip count. When the number of choices is an odd number, exclude the median choice and apply the same procedure. Next, sort H from greatest to smallest choice, and compute the distance from m for each choice. Sort L from smallest to greatest choice, and compute the distance for m from each choice. Finally, compute the difference in distances between each pair of median-adjusted scores in H and L. Under the null hypothesis of symmetry around m, the expected distance-difference between each pair of



(a) L1 (b) L2 (c) L3

(d) L4 (e) L5 (f) L6

(g) L7 (h) L8 (i) L9

Figure 3: **Treatment 1 CEPs vs. urn-filling choices**

choices is zero. However, these distance-differences are significantly different, well below the 1% confidence level for all small and large urns under the one-sample sign test.[13]

The filling choices for urns D.I and D.II share a similar pattern. The mean and the median urn-filling choices of winning chips in D.I are 4.2 and 5.0, respectively, where the range of feasible chips is 0-6. In D.II the mean and median are 5.1 and 5.0, respectively, where the range of feasible chips is 3-9. The median chip-filling choice in D.I and D.II is significantly different from the midpoint

[13]This test of symmetry is related to the skewness test in Doksum et al. (1977). However, the latter test requires a continuous distribution of choices, which does not conform well to the strong modes apparent in our data, nor to the severely restricted range of choices in the small urns.

m to many significant digits. Therefore, for both D.I and D.II, the null hypothesis of symmetry is rejected at the 1% confidence level.

The filling choices for urn E stands in sharp contrast to the choices in the other urns. There are seven possible urn-filling choices in E: a number of blue chips between 0 and 6. However, in urn E, 54% of the fill choices are exactly equal to the midpoint m. Further, nearly 80% of the choices are within one chip of m (i.e., either 2, 3, or 4 blue chips). The median choice of blue chips in E is not significantly different from 3, and although the skewness test is significant (p-value .012), skewness is small (Pearson's skewness coefficient is .188).

Overall, apart from urn E, Hypothesis 6 is robustly rejected. Accordingly, we conclude that urn-filling choices are not symmetrically distributed about m, but rather are skewed away from the midpoint and biased toward the interior boundary, b, of the winning range of probabilities (chips). In fact, the interior boundary itself is a powerful attractor: 38% of choices across L1, L2, L8, and L9 are exactly equal to this boundary (in L1 and L9 there are 21 possible choices, and in L2 and L8 there are 41 possible choices). In L3, L4, L6, and L7, where it is feasible to choose an equal number of both chip colors, either choosing this equal split or choosing the interior boundary b represent 40% of all choices. These two choices are clear attractors (in L3 and L7 there are 61 possible choices, and in L4 and L6 there are 81 possible choices). Small urns exhibit a similar pattern, as can be observed in Figure 5 in the Appendix.

Finding 4 ***Urn-filling choices pattern:*** *Contradicting Hypothesis 6, the numbers of chips placed in the urns are not symmetrically distributed about the midpoint of the allowable range, m. Instead, urn-filling choices are closer to the interior boundary, b, of the range of winning chips and are skewed away from the exterior boundary. The exception is urn E, where choices are tightly and symmetrically distributed about m.*

Recall that urn-filling subjects are informed neither about the winning color nor that other subjects will bet based on the urns they fill. While it is plausible that college student subjects might infer that to help other subjects they should create more equally distributed colors in the urns (subject to feasibility), it seems quite unlikely that 9-10 year-old students would attempt to anticipate how the urns they fill will be subsequently used. The distributions of urn-filling choices of these fourth-grade students are quite similar to the distributions of college student urn-filling choices. Thus, in most urns, there appears to be an innate attraction to place a number of chips between the midpoint m and the interior constraint b placed on chip colors. Particularly salient is the choice of b itself, as well as the choice of an equal number of both chips (the latter choice is only feasible in urns L3-L7).

IV.5 Revealed attitude toward ambiguity: is it a matter of perception?

Could the revealed attitude toward ambiguity characterized in Findings 1 and 2 be driven, to some extent, by the same mechanism underlying the urn-filling choices described in Finding 4? In a neural imaging study, Huettel et al. (2006) find that decision times are significantly *faster* when choosing between an ambiguously distributed outcome and a certain outcome, and *slower* when choosing between two risk-only outcomes. They also find that an external (non-neural) measure of behavioral impulsiveness predicts activation in the specific brain region associated with preferences for ambiguity, but *not* in the region associated with preferences for risk. Combined with the imaging data, the authors conclude that "decision making under ambiguity does not represent a special, more complex case of risky decision making; instead, these two forms of uncertainty are supported by distinct [neural] mechanisms." Thus, it appears to be the case that choices related to ambiguity are less deliberative than those related to risk, despite the (objectively) increased complexity of the decision problem. It seems plausible that, at least for some subjects, choices are influenced by an instinctive reaction to salient features of the stated range of probabilities, the same features that drive urn-filling choices.[14]

The impulsive nature of choices in the face of ambiguity helps rationalize the evidence that the source of ambiguity has little impact on the distribution of choices (Oechssler and Roomets, 2015), even when one of those sources is a uniform randomization device, as in Halevy (2007) and in our own Finding 3. Impulsivity is also consistent with the considerable extent to which individual choices between ambiguous and risk-only alternatives are dynamically inconsistent (Charness et al., 2013; Duersch et al., 2017; Bleichrodt et al., 2018). It seems plausible that some subjects make probabilistic choices based on "perception draws," weighted more heavily near cognitive reference

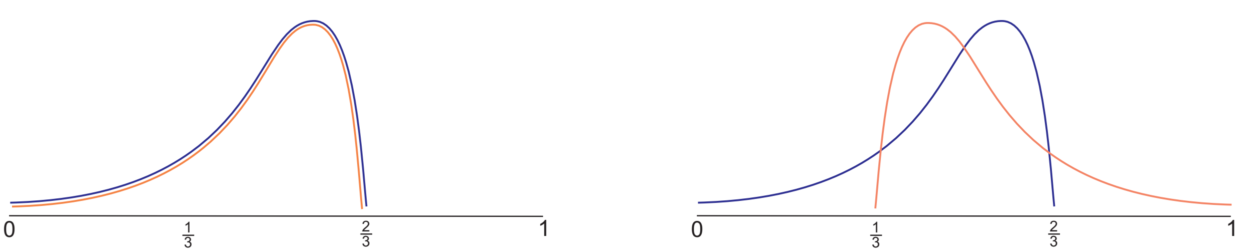
[14] Interestingly, subjects who reveal ambiguity-neutral attitudes, along with those who deviate the most from ambiguity neutrality, make the quickest decisions, while those who reveal themselves to have intermediate ambiguity attitudes have the slowest response times (these data, a subset of which are used in Stanton et al., 2011, are from private correspondence with Scott Huettel). This result is consistent with the possibility that those who reveal the most extreme attitudes act the most impulsively, those who are ambiguity neutral have an easy decision problem, while those with moderate attitudes exert effort to coordinate their instinctive reaction with the natural economic benchmark of midpoint m.

points (Rosch, 1975; Holyoak and Mah, 1982). The evidence in Hong et al. (2018) is consistent with this conclusion. They report that high-comprehension subjects reveal consistent ambiguity preferences, while low-comprehension subjects randomize. We suggest that this randomization is influenced by an underlying attraction to a particular subset of the feasible probability range, which we show to also be salient for urn fillers.

Of course, the fact that the pattern of urn-filling choices is similar to the pattern of CEPs does not necessarily imply that this shared pattern is caused by the same underlying perceptual process. Next, we provide indirect support for the conjecture that this pattern is caused by the same underlying perceptual process by studying the exogenous variation in perceptual salience through the lens of two parallel frames of the Ellsberg (1961) decision problems.

IV.6 Spotlighting the Ellsberg questions

In the Ellsberg three-color frame, the addition of a third color may break the salience of the interior boundary of possible probabilities b (expressed as a fraction of total chips). To illustrate this, suppose that blue is the winning color. In the two-color urn 1 in D.I, the fraction of blue chips is $\left[0, \frac{2}{3}\right]$, and the fraction of the remaining orange chips is $\left[\frac{1}{3}, 1\right]$. Thus, the interior boundary for blue chips is $b = \frac{2}{3}$, and for orange chips is $b = \frac{1}{3}$. As the fraction of blue chips increases toward its maximum, $\frac{2}{3}$, the fraction of orange chips necessarily decreases toward its minimum, $\frac{1}{3}$, such that b is approachable in both colors. In the three-color urn, due to the presence of $\frac{1}{3}$ gray chips, the fraction of both blue and orange chips is between 0 and $\frac{2}{3}$. Thus, the salient fraction for both blue and orange chips is the interior boundary $b = \frac{2}{3}$. However, implementing both simultaneously is not feasible. Figure 4 is a visual representation of the perceptual salience of the possible blue and orange chip fractions in E.I and D.I.



(a) Three-color choice (E.I) (b) Two-urn choice (D.I)

Figure 4: **Perceptual salience**

The three-color urn breaks the complement to event blue into two events, orange and gray, thereby breaking the complementarity of the interior boundary of blue and orange. Whereas urn 1 in the two-color experiment maintains the complementarity of the interior boundary of blue and orange, allowing the perceptual salience of both ranges to impact choices simultaneously. This difference provides a plausible explanation for the revealed preference reversal between E.I and D.I.

This conjecture about the break in the complementarity of the interior boundary is based upon reviewing the urn-filling and incentivized choices in Treatments 1-5. To test this conjecture out of sample, we introduce two-urn decision problem D.II into Treatments 6 and 7. As reported in Table 5, in the three-color decision problem E.II, only 33% of subjects prefer the ambiguous alternative to the risk-only alternative. Thus, subjects' revealed preferences are consistent with ambiguity aversion for likely gains, as expected. To test the conjecture that a two-urn frame induces a change in the perceptual salience of the underlying distributions of winning and losing, in Treatment 6, we implement the dual D.II to E.II and test whether subjects in this re-framing are significantly more ambiguity averse than subjects in E.II. Table 5 reports that in D.II a mere 15% of subjects choose the ambiguous urn over the risk-only urn, a highly significant difference from the 33% who choose the ambiguous alternative in E.II (the Fisher's Exact test p-value is zero to many significant digits). These findings are consistent with the conjecture that the change in perceptual salience of the underlying distributions of winning and losing is causing a change in the behavior of DMs.

Note that breaking the salience of interior boundary b for the two ambiguous colors does not move DMs to treat all colors as equally likely in both decision problems. In E.II, while the change in frame to three colors does move choices in the direction of uniformly distributed states (Fox and Clemen, 2005), the bias toward uniformly distributed consequences is still maintained, albeit to a lesser extent. In D.II there is a one-to-one mapping between states and outcomes (payoffs), while in E.II there is many-to-one relation. The latter case generates a tension between perceptual salience in the direction of uniformly distributed states and perceptual salience in the direction of uniformly distributed payoffs. This tension mitigates the effect of perceptual salience toward uniformity in state probabilities.

We also note that support theory (Tversky and Koehler, 1994; Rottenstreich and Tversky, 1997), which holds that the judged probability of an event generally increases when its description is unpacked into disjoint components, cannot be tested with our design. Support theory implies that subjects would prefer the ambiguous alternative in D.I over the ambiguous alternative in E.I (where the losing event has been unpacked), and they would prefer the ambiguous alternative in E.II (where the winning event has been unpacked) over the ambiguous alternative in D.II. But such choices are not elicited in our experiment.

We summarize the findings for Ellsberg-urn choices, confirming Hypothesis 7, as follows.

Finding 5 ***Ellsberg ordering:*** *The fraction of subjects who choose the ambiguous alternative over the risk-only alternative in the Ellsberg questions aligns in the following order, from smallest to greatest: D.II, E.II, E.I, E1.I, E2.I, D.I, supporting Hypothesis 7. Of particular importance, the difference between E.I and its dual D.I, and between E.II and its dual D.II, are highly statistically significant.*

IV.7 Subtreatment effects

To test for subtreatment effects on CEPs, we implement the Kruskal-Wallis H test for each urn within each treatment. Overall, of 36 large and small urns in Treatments 1 and 2, only in two urns do CEPs differ significantly by color subtreatment at even the 10% confidence level. That is, the choices in all other 34 urns do not significantly depend on whether blue or orange is the winning color. Choices also do not significantly depend on whether small or large urns are presented first: only in 3 of 36 urns are the choices different by presentation order at the 10% confidence level. Finally, choices differ significantly by subtreatment in only 1 of 36 urns at the 10% confidence level. Thus, overall, we observe subtreatment effects only in about 5% of the urns at the 10% confidence level, suggesting there are no systematic subtreatment effects.

In Treatment 6, we do identify a significant subtreatment effect in three out of the nine small urn choice problems. Subjects who first viewed the two three-color Ellsberg decision problems, then the nine small urns, then the two two-urn Ellsberg dual decision problems (subtreatment ESD) tend to be more averse to ambiguity across urns S1-S9 than subjects who first view the two two-urn Ellsberg dual decision problems, then the small urns, then the three-color Ellsberg decision problems (subtreatment DSE). The difference is significant at the 5% confidence level for S1 and S6 (p-values of .033 and .027, respectively), and at the 10% confidence level for S7 (p-value .084). For these urns there is a 11-19 percentage point increase in the proportion of subjects who prefer the risk-only urn to the ambiguous urn. The difference for the remaining six small urn decision problems is 0 to 10 percentage points, with p-values ranging from .16 to .93. In Treatment 6, we do not identify any statistically significant differences in the four Ellsberg decision problems themselves by subtreatment (p-values between .13 and .69). The characterization of choices reported in Findings 1 and 2 are consistent when analyzed by subtreatment. Therefore, subtreatments have negligible impact on our results.

Finding 6 ***Subtreatment effects:*** *There are no significant winning color, question order (or both) effects in Treatments 1 and 2. In Treatment 6, there is a persistent increase in revealed aversion to ambiguity in S1-S9 when moving from subtreatment DES to ESD. However, this difference is significant at the 10% confidence level for only three of the nine decision problems, and there is no significant order effect for the Ellsberg questions themselves.*

IV.8 Urn size and endowment effects

Next, we test for the effect of urn size on elicited CEPs, and the effect of endowment on elicited CEPs in MPL. For each Treatment 1 and 2, we compare the choices in each large urn to the choices in its corresponding small urn (e.g., we compare L1 to S1, L2 to S2, etc.). In particular, we test the null hypothesis that the CEPs elicited from the large and small urns are drawn from the same population distribution, using the Mann-Whitney Rank Sum test. We do not reject the null hypothesis for any winning probability range (urn comparison); the smallest p-value over all 18 tests is .328. Thus, we conclude that subjects do not record significantly different CEPs for large relative to small urns.

Finding 7 ***Urn size effect:*** *There is no significant urn size effect in Treatments 1 and 2.*

We expect subjects "endowed" with an ambiguous bet (facing a fixed ambiguous bet relative to a list of risk-only bets) to reveal a stronger preference for ambiguity than when facing a fixed risk-only bet relative to a list of ambiguous bets. In Treatment 3, there are three binary choices that involve the same pair of ambiguous and risk-only bets, but with opposing endowments. Comparing the fraction of subjects who prefer the risk-only urn to the ambiguous urn, split by endowment type, the difference is insignificant: Fisher's Exact test p-values for L3, L5, and L7 are 1, .233, and .815, respectively. Thus, we do not identify an endowment effect for ambiguity relative to risk-only.

This finding is surprising, given the strong endowment effect identified for a fixed risk-only lottery relative to a list of risk-only lotteries that vary in probabilities (Sprenger, 2015). This finding provides further evidence that ambiguity is evaluated by processes quite different from those used to evaluate risk.

Finding 8 ***Endowment effect:*** *There is no significant endowment effect in Treatment 3.*

V Conclusion

This paper identifies an important insight, formerly "hidden" between the two alternative elicitation frames of the famous Ellsberg (1961) Paradox: individuals' revealed preferences for ambiguous versus risk-only bets depend on more than the respective unknown probabilities of the payoffs. In particular, the three-color one-urn frame induces less pronounced preferences, while the dual two-color two-urn representation induces far stronger revealed preferences: increased ambiguity seeking for unlikely gains, and increased ambiguity aversion for likely gains. In fact, the difference between the responses for the Ellsberg three-color one-urn choices more than **triples** when moving to the dual two-color two-urn frame.

We attribute this change in revealed preferences to non-uniform perceptual salience of the underlying ranges of probabilistic states of nature. The three-color one-urn and two-color two-urn frames differentially impact revealed preferences for ambiguity through subjective perceived probabilities. In particular, in the two-color two-urn frame, when facing state probabilities that coincide with payoff probabilities, individuals are drawn toward the interior relative to exterior bounds of the range of winning probabilities, consistent with a bias toward a uniform distribution of outcomes (coinciding states and payoffs). However, in the three-color one-urn frame, when a single payoff is associated with two states, an induced bias towards a uniform distribution of states mitigates the bias toward a uniform distribution of payoffs. We base this insight on the identification of a similar pattern, but in exaggerated fashion, for individuals who fill urns with no financial stake in their choices, including children as young as nine years old.

Daniel Kahneman asserts that, "In visual perception, you have a process that suppresses ambiguity, so that a single interpretation is chosen, and you're not aware of the ambiguity."[15] We

[15] In a January 13, 2017 interview with Drake Baer for The Cut.

propose that a related perceptual process focuses human attention on particular features of probability ranges, subconsciously influencing perceived probabilities and ensuing choices. This notion is supported by the similarity of urn-filling choices between fourth graders and college students, who would appear to have very few conscious reasons to fill urns in similar proportions, as well as by the high correlation between the bias in the choices of urn fillers and the perceived probabilities of incentivized decision makers, across different frames.

Our findings are consistent with prior studies which indicate that subjects facing decision problems featuring ambiguity tend to act impulsively (Huettel et al., 2006), inconsistently (Charness et al., 2013; Duersch et al., 2017; Bleichrodt et al., 2018), and with substantial randomization by low-comprehension subjects (Hong et al., 2018). Our findings may help explain why individuals tend to treat compound lotteries as ambiguous (Armantier and Treich, 2015), and why individuals who exhibit non-neutral attitudes toward ambiguity tend to reduce compound lotteries incorrectly (Halevy, 2007; Abdellaoui et al., 2015). We believe that the current study provides a foundation from which to explore the extent to which perceptual salience guides choices.

Our finding that perceived probabilities and consequent decisions share robust characteristics that can be managed (manipulated) by designing a particular partition of the state space into events has many important theoretical and practical implications. For example, credit default swaps induce a particular partition of the state space in which all solvency states are unionized, affecting perceived probabilities and therefore spreads (Augustin and Izhakian, 2019). Bonds, derivative assets, and capital structure decisions are affected similarly. Generally, incentivizing contracts can be designed to induce a particular partition of the state space into events faced by each counterpart, thereby managing perceived ambiguity and probabilities and improving contract efficiency (Izhakian and Zender, 2018).

With regard to medical decisions, obesity—a critical problem in public health—for example, increases the likelihood of premature death due to both cardiovascular and non-cardiovascular complications. Thus patients may be more likely to respond affirmatively to mortality risks if these risks are itemized rather than bundled. Similarly, an individual may tend to purchase insurance when consequences with ambiguous likelihoods are itemized, which could improve the uptake of partial insurance programs (e.g., drought insurance) that appear to be under-utilized (Bryan, 2019). The same reasoning applies to influence policy makers who consider the consequences of climate change, water shortages, and urban development.

References

Abdellaoui, M., A. Baillon, L. Placido, and P. P. Wakker (2011) "The Rich Domain of Uncertainty: source Functions and Their Experimental Implementation," *American Economic Review*, Vol. 101, No. 2, pp. 695–723.

Abdellaoui, M., P. Klibanoff, and L. Placido (2015) "Experiments on compound risk in relation to simple risk and to ambiguity," *Management Science*, Vol. 61, No. 6, pp. 1306–1322.

Armantier, O. and N. Treich (2015) "The rich domain of risk," *Management Science*, Vol. 62, No. 7, pp. 1954–1969.

Augustin, P. and Y. Izhakian (2019) "Ambiguity, Volatility and Credit Risk," *The Review of Financial Studies*, Forthcoming.

Becker, G. M., M. H. DeGroot, and J. Marschak (1964) "Measuring utility by a single-response sequential method," *Behavioral Science*, Vol. 9, No. 3, pp. 226–232.

Binmore, K., L. Stewart, and A. Voorhoeve (2012) "How much ambiguity aversion? Finding indifferences between Ellsberg's risky and ambiguous bets," *Journal of Risk and Uncertainty*, Vol. 45, No. 3, pp. 215–238.

Bleichrodt, H., J. Eichberger, S. Grant, D. Kelsey, and C. Li (2018) "A test of dynamic consistency and consequentialism in the presence of ambiguity."

Brenner, M. and Y. Izhakian (2018) "Asset Prices and Ambiguity: Empirical Evidence," *Journal of Financial Economics*, Vol. 130, pp. 503–531.

Bryan, G. (2019) "Ambiguity Aversion Decreases the Impact of Partial Insurance: Evidence from African Farmers," *Journal of the European Economic Association*, Forthcoming.

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

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

Crockett, S. and B. E. Crockett (2019) "Endowments and risky choice," *Journal of Economic Behavior and Organization*, Vol. 159, pp. 344–354.

Curley, S. P. and J. F. Yates (1985) "The center and range of the probability interval as factors affecting ambiguity preferences," *Organizational Behavior and Human Decision Processes*.

——— (1989) "An empirical evaluation of descriptive models of ambiguity reactions in choice situations," *Journal of Mathematical Psychology*, Vol. 33, No. 4, pp. 397–427.

Dhillon, A. and J.-F. Mertens (1999) "Relative utilitarianism," *Econometrica*, Vol. 67, pp. 471–498.

Di Mauro, C. and A. Maffioletti (2004) "Attitudes to risk and attitudes to uncertainty: experimental evidence," *Applied Economics*, Vol. 36, No. 4, pp. 357–372.

Dimmock, S. G., R. Kouwenberg, and P. P. Wakker (2016) "Ambiguity attitudes in a large representative sample," *Management Science*, Vol. 62, No. 5, pp. 1363–1380.

Doksum, K., G. Fenstad, and R. Aaberge (1977) "Plots and tests for symmetry," *Biometrika*, Vol. 64, No. 3, pp. 473–487.

Duersch, P., D. Römer, and B. Roth (2017) "Intertemporal stability of uncertainty preferences," *Journal of Economic Psychology*, Vol. 60, pp. 7–20.

Ellsberg, D. (1961) "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, Vol. 75, No. 4, pp. 643–669.

Filiz-Ozbay, E., H. Gulen, Y. Masatlioglu, and E. Y.Ozbay (2018) "Size matters under ambiguity," mimeo.

Fox, C. R. and R. T. Clemen (2005) "Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior," *Management Science*, Vol. 51, No. 9, pp. 1417–1432.

Freeman, D., Y. Halevy, and T. Kneeland (2018) "Eliciting risk preferences using choice lists," *Quantitative Economics*.

Gonzalez, R. and G. Wu (1999) "On the shape of the probability weighting function," *Cognitive psychology*, Vol. 38, No. 1, pp. 129–166.

Grether, D. (1981) "Financial incentive effects and individual decision-making,"Technical Report 401, California Institute of Technology Social Science Working Paper.

Halevy, Y. (2007) "Ellsberg revisited: an experimental study," *Econometrica*, Vol. 75, No. 2, pp. 503–536.

Harrsion, G. W. and E. Rutström (2008) "Risk aversion in the laboratory," in G. W. Harrsion and J. C. Cox eds. *Risk Aversion in Experiments*: Bingley, UK: Emerald, Research in Experiment Economics, Volume 12.

Holt, C. A., S. K. Laury et al. (2002) "Risk aversion and incentive effects," *American economic review*, Vol. 92, No. 5, pp. 1644–1655.

Holyoak, K. J. and W. A. Mah (1982) "Cognitive reference points in judgements of symbolic magnitude," *Cognitive Psychology*, Vol. 14, pp. 328–352.

Hong, C. S., M. Ratchford, and J. S. Sagi (2018) "You need to recognize ambiguity to avoid it," *The Economic Journal*, Vol. 128, No. 614, pp. 2480–2506.

Huettel, S. A., C. J. Stowe, E. M. Gordon, B. T. Warner, and M. L. Platt (2006) "Neural Signatures of Economic Preferences for Risk and Ambiguity," *Neuron*, Vol. 49, No. 5, pp. 765–775.

Izhakian, Y. (2017) "Expected Utility with Uncertain Probabilities Theory," *Journal of Mathematical Economics*, Vol. 69, pp. 91–103.

——— (2018) "A Theoretical Foundation of Ambiguity Measurement," *SSRN eLibrary*, 1332973.

Izhakian, Y. and J. Zender (2018) "Ambiguity and Beliefs in The Principal-Agent Model," *SSRN eLibrary*, 3066761.

Karni, E. (2009) "A mechansim for eliciting probabilities," *Econometrica*, Vol. 77, No. 2, pp. 603–606.

Kocher, M. G., A. M. Lahno, and S. T. Trautmann (2018) "Ambiguity aversion is not universal," *European Economic Review*, Vol. 101, pp. 268–283.

Li, X. and C. Camerer (2019) "Using visual salience in empirical game theory," mimeo.

Mukhtar, A. (1977) "Probability and utility estimates for racetrack betting," *Journal of Political Economy*, Vol. 85, pp. 803–815.

Oechssler, J. and A. Roomets (2015) "A test of mechanical ambiguity," *Journal of Economic Behavior & Organization*, Vol. 119, pp. 153–162.

Plott, C. R. and K. Zeiler (2005) "The willingness to pay-willingness to accept gap, the 'endowment effect,' subject misconceptions, and experimental procedures for eliciting valuations," *The American Economic Review*, Vol. 95, No. 3, pp. 530–545.

Pulford, B. D. and A. M. Colman (2008) "Size doesn't really matter," *Experimental Psychology*, Vol. 55, No. 1, pp. 31–37.

Rosch, E. (1975) "Cognitive reference points," *Cognitive Psychology*, Vol. 7, No. 4, pp. 532–547.

Rottenstreich, Y. and A. Tversky (1997) "Unpacking, Repacking, and Anchoring: Advances in Support Theory," *Psychological Review*, Vol. 104, No. 2, pp. 406–415.

Slusser, E. and H. Barth (2017) "Intuitive proportion judgment in number-line estimation: Converging evidence from multiple tasks," *Journal of Experimental Child Psychology*, Vol. 162, pp. 181–198.

Spence, I. (1990) "Visual psychophysics of simple graphical elements," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 16, No. 4, pp. 683–692.

Sprenger, C. (2015) "An endowment effect for risk: Experimental tests of stochastic reference points," *Journal of Political Economy*, Vol. 123, No. 6, pp. 1456–1499.

Stahl, D. O. (2014) "Heterogeneity of ambiguity preferences," *Review of Economics and Statistics*, Vol. 96, No. 4, pp. 609–617.

Stanton, S. J., O. A. Mullette-Gillman, R. E. McLaurin, C. M. Kuhn, K. S. LaBar, M. L. Platt, and S. A. Huettel (2011) "Low- and high-testosterone individuals exhibit decreased aversion to economic risk," *Psychological Science*, Vol. 22, No. 4, pp. 447–453.

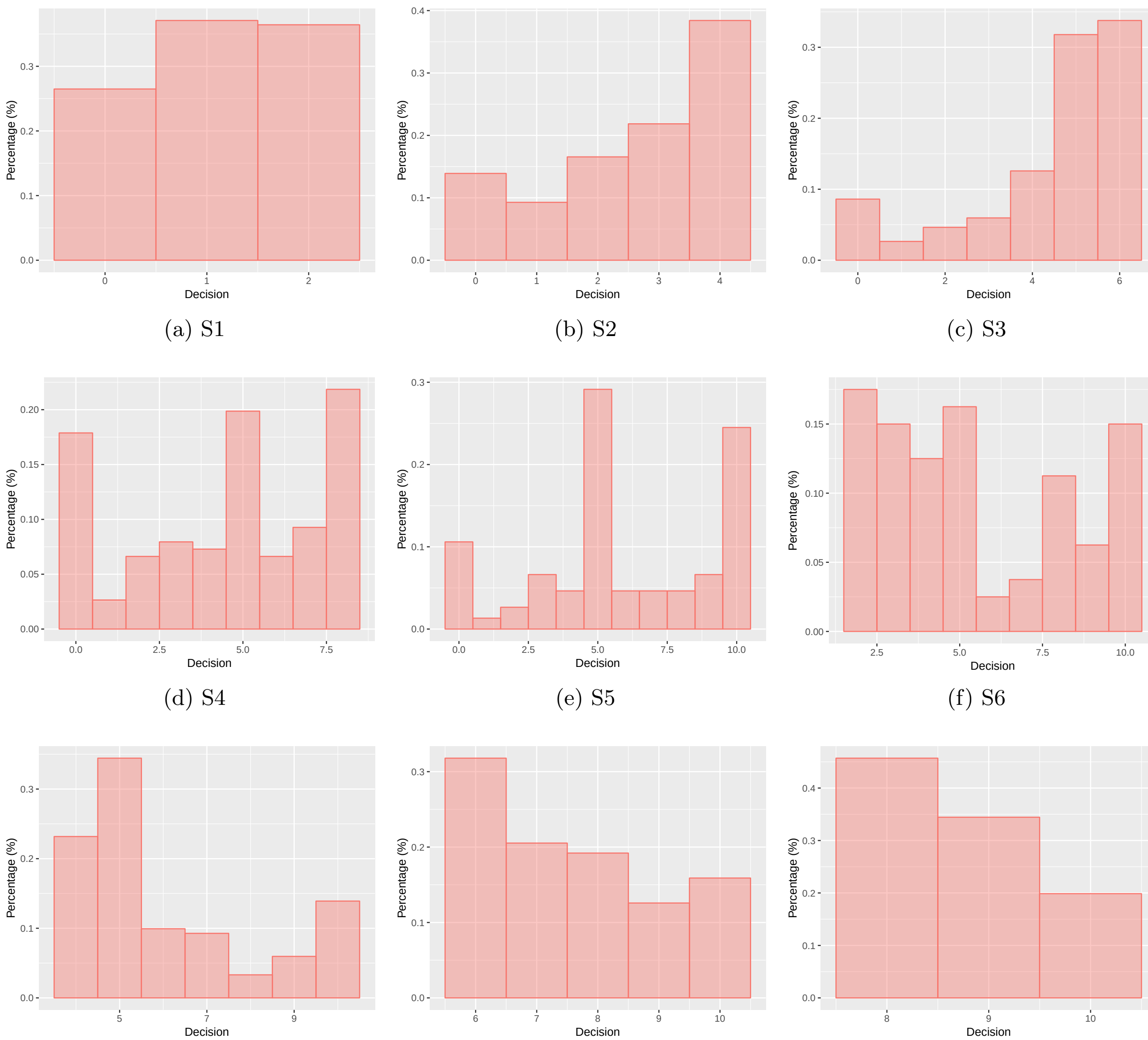
Trautmann, S. T. and G. van de Kuilen (2015) "Ambiguity attitudes," in G. Keren and G. Wu eds. *The Wiley Blackwell Handbook of Judgement and Decision Making, II*: John Wiley & Sons, Ltd.

Tversky, A. and D. Kahneman (1992) "Advances in prospect theory: cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, Vol. 5, No. 4, pp. 297–323.

Tversky, A. and D. J. Koehler (1994) "Support Theory: A Nonextensional Representation of Subjective Probability," *Psychological Review*, Vol. 101, pp. 547–567.

Wertheimer, M. (1938) "Numbers and numerical concepts in primitive peoples," in W. D. Ellis ed. *A source book of Gestalt psychology*: Harcourt, Brace, and Co.

Appendix



(a) S1 (b) S2 (c) S3

(d) S4 (e) S5 (f) S6

(g) S7 (h) S8 (i) S9

Figure 5: **Small urn fill choices**