A test of mechanical ambiguity

Jörg Oechssler\textsuperscript{a,}\textsuperscript{*}, Alex Roomets\textsuperscript{b}

\textsuperscript{a} Department of Economics, University of Heidelberg, Germany
\textsuperscript{b} Department of Economics, Franklin & Marshall College, United States

\textbf{A R T I C L E   I N F O}

Article history:
Received 22 July 2014
Received in revised form 28 July 2015
Accepted 7 August 2015
Available online 20 August 2015

JEL classification:
C91
D81

Keywords:
Ambiguity aversion
Uncertainty
Experiment
Ellsberg

\textbf{A B S T R A C T}

We implement an experiment to elicit subjects’ ambiguity attitudes in the spirit of Ellsberg’s three-color urn. The procedure includes three design elements that (together) have not been featured in similar experiments: strict ambiguity preferences, a single decision, and a mechanical randomization device with an unknown distribution (to both subjects and experimenters). We use this device in order to eliminate possible “strategic” ambiguity related to subjects’ beliefs about the experimenters’ motivations. In addition, we survey 40 experimental studies on Ellsberg’s two- and three-color problems, and find that, on average, slightly more than half of subjects are classified as ambiguity averse. Our results, with our new design, fall on the low end of the range of results in the surveyed studies, and are comparable to a control test where “strategic” ambiguity was induced.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

There are many situations where decision makers must contend with natural phenomena which are not well understood. Take for example the first man to ever eat a mushroom. We might now know if the mushroom in question is toxic, and, if so, the probability of various reactions. This man, on the other hand, would have known nothing of the sort. Similarly, consider the first men to plant non-native crops in various parts of the world, chemists experimenting with new compounds, or early explorers on uncharted oceans. Each deals with what economists generally refer to as ambiguity. Further, the ambiguity in these instances stems primarily from acts of nature, which have no discernible strategic underpinnings.

In order to better understand the decision-making process in these situations we propose and conduct a lab experiment. While other experiments have looked at ambiguity, the source nearly always stems from some conscious decision maker. We instead take steps to ensure that the ambiguity stems from a mechanical process, which not even we as experimenters fully understand.\textsuperscript{1} In order to maintain comparability to the existing literature in terms of results, we utilize a three-color Ellsberg urn with one risky and two ambiguous colors.

\textsuperscript{*} We thank Aurelien Baillon, Peter Duersch, Jürgen Eichberger, P.J. Healy, Stefan Trautmann, Peter Wakker and participants of talks at the Universities of Heidelberg and Rotterdam and the 2012 ESA Conference in Tucson for many useful discussions. The Co-Editor, the Associate Editor and two referees added valuable comments. The University of Heidelberg provided financial support.

\textsuperscript{*} Corresponding author. Tel.: +49 6221 54 3548; fax: +49 6221 54 3630.

E-mail addresses: oechssler@uni-hd.de (J. Oechssler), roomets@gmail.com (A. Roomets).

\textsuperscript{1} Natural processes have been used before to study ambiguity, but in ways that could be manipulated by the experimenter. We discuss these studies further in Section 3.1.
In order to gain clean evidence on the effects of ambiguity from mechanical phenomena, we incorporate three design features that, in combination, have not before been implemented. The first feature is to differentiate between mechanical and strategic sources of ambiguity by constructing a mechanical randomization device with an unknown distribution (to both subjects and experimenters). The second is to eliminate the possibility of ambiguity “hedging” by sophisticated subjects. The last is to avoid mischaracterization due to indifference.

Upon implementing our design, we find that subjects are averse, in roughly equal proportions, to ambiguity from both sources. With our design this proportion seems to be around 40%. We also find that roughly 25% of subjects make ambiguity seeking choices, though further investigation of their reasoning allows for a different interpretation of this statistic. We additionally find high correlation between subjects’ reasoning and their choices. In particular, subjects who emphasized the safety of the risky option (as compared to an ambiguous option) all chose the risky option. Meanwhile, subjects whose reasoning was congruent with the principle of insufficient reason were 40% more likely to choose the option suggested by that principle.

To put our results into perspective, in the next section we survey 40 experimental studies on the Ellsberg urn problem and find that on average slightly more than half of subjects were classified as ambiguity averse. With our new design features, we find that probably slightly less than half should be classified as ambiguity averse. Section 3 discusses in detail the choices we made with respect to the experimental design and procedures. Results are analyzed and discussed in Section 4. Finally, we close with a brief discussion of the implications of our findings in Section 5.

2. A survey of the experimental literature

Since Ellsberg’s (1961) famous thought experiment there have been numerous laboratory experiments that implemented either the 2-color or the 3-color Ellsberg urn. Camerer and Weber (1992) provide a survey of the early literature. Since in recent years the experimental literature seems to have exploded, an updated survey may be valuable. Trautmann and van de Kuijen (2014) provide an excellent survey focussing on the premium that subjects are willing to pay in order to avoid ambiguity. In Table 1 we provide a survey of 39 experimental studies focussing on the percentage of subjects that can be classified as ambiguity averse.

We have collected all studies we could find that implemented a classical Ellsberg urn experiment and where we were able to infer the percentage of ambiguity averse subjects from data presented in the paper (column five in Table 1). Whenever possible, we chose the treatment that was closest to the classical Ellsberg experiment.

The second column of Table 1 notes whether it was the 2-color or the 3-color version. Column three indicates whether the experiment was designed to elicit a strict preference regarding ambiguity (e.g. by eliciting an ambiguity premium or by paying a higher price for the ambiguous lottery). The fourth column indicates whether the experiment involved multiple decisions for the same subject (which often allows for some form of ambiguity hedging).

Table 1 shows that the range of the percentage of ambiguity averse subjects is large (from 8% to 93%). If we naively take a simple (unweighted) average of (the midpoints of the ranges) of all studies, we obtain a mean of 56.6 and a median of 58.5. If we discard the five most extreme studies on either side, the mean is almost unchanged at 57.1. If we only include studies that check for a strict preference toward ambiguity, we obtain an average of 51.2%. Thus, overall, a quite consistent picture emerges: Slightly more than half of subjects seem to be ambiguity averse.2

3. Experimental design

In the following three subsections we will discuss the three crucial features of our experimental design.

3.1. Strategic vs. mechanical ambiguity

Ambiguity can have two different types depending on its source. One type, which we shall call mechanical ambiguity, is the ambiguity about the distribution of a variable determined by some mechanical process. The second type, which we shall call strategic ambiguity, is the ambiguity about the distribution of a random variable determined by the conscious decision of some other agent.3

In a standard Ellsberg-urn experiment, subjects may perceive their decision problem as a game between whoever fills the urn and themselves. For example, subjects may believe that the experimenter is trying to trick them in order to protect his research budget.4 Therefore, it can be said that these experiments are technically a case of strategic ambiguity. While strategic ambiguity is certainly important to study, as much ambiguity in life comes from conscious decision, it is not clear that results in these experiments can be applied to cases where ambiguity comes from unknown mechanical processes.

---

2 The remaining subjects would be classified as either ambiguity loving or ambiguity neutral.

3 Thus, the two types of ambiguity differ in a manner similar to the difference between nodes of game trees where players move, and nodes where nature moves.

4 This concern has been raised in a number of experimental papers (e.g. Keren and Gerritsen, 1999; Charness et al., 2013; Dominiak and Duersch, 2012). See Trautmann et al. (2008) for an interesting approach countering this concern.
Table 1
Previous Ellsberg experiments.

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of colors</th>
<th>Strict pref</th>
<th>Mult. dec</th>
<th>% Averse</th>
<th>Other comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akay et al. (2012)</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Becker and Brownson (1964)</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Binmore et al. (2012)</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>45–64</td>
<td></td>
</tr>
<tr>
<td>Butler et al. (2014)</td>
<td>2</td>
<td>Yes</td>
<td>No</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Butler et al. (2013)</td>
<td>2</td>
<td>No</td>
<td>Yes</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Czellar and Riedl (2010)</td>
<td>2</td>
<td>No</td>
<td>Yes</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Charness et al. (2013)</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>8–24</td>
<td>Versions 1 and 2</td>
</tr>
<tr>
<td>Chew et al. (2012)</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Chew et al. (2013)</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>62</td>
<td>50 vs. [0,100]</td>
</tr>
<tr>
<td>Cohen et al. (1985)</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>Cohen et al. (1987)</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>Cooper and Johnson (2013)</td>
<td>2</td>
<td>No</td>
<td>No</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Curley et al. (1986)</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>65</td>
<td>Experiment I</td>
</tr>
<tr>
<td>Dimmock et al. (2015)</td>
<td>2</td>
<td>No</td>
<td>Yes</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>Dimmock et al. (2013)</td>
<td>2</td>
<td>No</td>
<td>Yes</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>Dominik et al. (2012)</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>Dominik et al. and Duersch (2012)</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>23–48</td>
<td></td>
</tr>
<tr>
<td>Dominik et al. and Schnedler (2011)</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>55</td>
<td>All subjects</td>
</tr>
<tr>
<td>Eichberger et al. (2015)</td>
<td>2</td>
<td>No</td>
<td>No</td>
<td>63</td>
<td>Situation O</td>
</tr>
<tr>
<td>Einhorn and Hogarth (1986)</td>
<td>2</td>
<td>No</td>
<td>Yes</td>
<td>47</td>
<td>MBA students</td>
</tr>
<tr>
<td>Eliazi and Orotleva (2015)</td>
<td>3</td>
<td>No</td>
<td>Yes</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>Fox and Tversky (1995)</td>
<td>3</td>
<td>No</td>
<td>Yes</td>
<td>62</td>
<td>Study 3</td>
</tr>
<tr>
<td>Gazzale et al. (2013)</td>
<td>2</td>
<td>No</td>
<td>Yes</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>Halevy (2007)</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Keck et al. (2012)</td>
<td>2</td>
<td>No</td>
<td>Yes</td>
<td>60</td>
<td>( p = 0.50, \delta = 0.50 )</td>
</tr>
<tr>
<td>Keren and Gerritsen (1999)</td>
<td>3</td>
<td>No</td>
<td>Yes</td>
<td>68–79</td>
<td></td>
</tr>
<tr>
<td>Kuehberger and Perner (2003)</td>
<td>2</td>
<td>No</td>
<td>Yes</td>
<td>75</td>
<td>Graphical estimate</td>
</tr>
<tr>
<td>Kuzmic (2013)</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>55</td>
<td>AA + AC</td>
</tr>
<tr>
<td>Lauriola and Levin (2001)</td>
<td>2</td>
<td>No</td>
<td>Yes</td>
<td>62–72</td>
<td>( p = 0.50 )</td>
</tr>
<tr>
<td>MacCrimmon and Larsson (1979)</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>79</td>
<td>70% reported in text?</td>
</tr>
<tr>
<td>Maffioletti and Santoni (2005)</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Ross et al. (2012)</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>27</td>
<td>Strong aversion</td>
</tr>
<tr>
<td>Rubinstein (2012)</td>
<td>3</td>
<td>No</td>
<td>Yes</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Slovic and Tversky (1974)</td>
<td>3</td>
<td>No</td>
<td>Yes</td>
<td>51–78</td>
<td>Averse reasoning</td>
</tr>
<tr>
<td>Stahl (2014)</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>35–39</td>
<td>125 or 65, Resp.</td>
</tr>
<tr>
<td>Trautmann et al. (2008)</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>50–79</td>
<td>Experiment I</td>
</tr>
<tr>
<td>Trautmann et al. (2011)</td>
<td>2</td>
<td>No</td>
<td>Yes</td>
<td>63</td>
<td>Experiment I</td>
</tr>
<tr>
<td>Vinogradov and Shadrina (2013)</td>
<td>2</td>
<td>No</td>
<td>Yes</td>
<td>59–74</td>
<td></td>
</tr>
<tr>
<td>Yates and Zukowski (1976)</td>
<td>2</td>
<td>No</td>
<td>Yes</td>
<td>78</td>
<td>( G ) vs. ( G^c )</td>
</tr>
</tbody>
</table>

Notes: Column “No. of colors” refers to whether it is a 2-color or 3-color Ellsberg experiment. Column “strict pref.” refers to whether the experiment can elicit strict preferences regarding ambiguity. Column “mult. dec.” refers to whether the experiment includes multiple decisions by one subject. Column “% averse” refers to the percentage of ambiguity averse subjects in the experiment. Note that the remaining subjects can be either ambiguity loving or neutral. Column “other comments” describes the particular treatment of the respective paper that we selected for comparison.

No monetary incentives.

On the one hand, mathematically, the source of ambiguity makes no difference. On the other, the psychological source of ambiguity aversion is not well understood, and there are reasons to believe that the source of ambiguity may affect someone’s aversion to it. In particular, failure in the face of strategic ambiguity may induce a feeling of inferiority that would not be present in the face of mechanical ambiguity. Similarly, success in the face of strategic ambiguity might induce analogous feelings of superiority.

We set out to answer the question of strategic versus mechanical ambiguity with an experiment using a new procedure for generating an ambiguous distribution that is mechanical in nature. The procedure involves creating an irregular Galton box. A Galton box (also called a bean machine) is a box with interleaved pins arranged in such a way that when a ball is dropped in the top of the box, it ricochets off the various pins and into a number of collection bins. In a regular Galton box, the pins are arranged in such a way that the balls fall into the various bins according to an approximately normal distribution. In an irregular Galton box, the distribution is irregular. Our goal is for this distribution to be ambiguous.

In order to test mechanical ambiguity, it is not enough to create an ambiguous distribution for subjects. We need the distribution to be ambiguous to the experimenters as well. To this end, we construct the frame of a Galton box before the experiment, but arrange the pins and complete construction in the lab, with the assistance of the subjects. Specifically, we have subjects hammer nails into a plywood board before revealing anything else about the experiment. We then attach the plywood board to the frame and reveal its purpose.

We reveal that we will drop a number of bouncy rubber super-balls through the Galton box, where they will make their way into black and white bins. For each ball that falls into a white (black) bin, we add a yellow (blue) marble to a
three–color Ellsberg urn. Using this procedure, we create a randomizing device with a distribution that is ambiguous to both experimenters and subjects.

Others have attempted to test some form of natural ambiguity before (Fox and Tversky, 1995; Kilka and Weber, 2001; Chow and Sarin, 2002; Maffioletti and Santoni, 2005; Keller et al., 2007; Abdellaoui et al., 2011; Füllbrunn et al., 2014; Butler et al., 2013, to name a few). These studies use natural phenomena, such as the weather, the stock market, or the content of M&M candy bags, to generate bets with ambiguous winning probabilities. Although when encountered in nature these phenomena would exhibit mechanical ambiguity, in the lab it is not as clear. If we expect a subject to be suspicious of an experimenter filling an urn, it is possible that similar suspicions may arise relating to the formulation of these bets. With respect to weather, for example, a plethora of data has been gathered and is easily searchable. Thus, the winning probability of a bet on the temperature in some city on a given day could be estimated fairly accurately beforehand. This allows the strategic motivations of the experimenter to come into play. A second reason why we prefer to use the Galton box is that making payments dependent on future events would necessarily delay payments thereby confounding issues of ambiguity with issues of discounting and time preferences.

### 3.2. Single vs. multiple decisions

We ask each participant to make just a single decision. Although this increases our cost for data, standard multiple–decision designs would not meet our needs. In particular, even a mechanism in which one decision is randomly selected for payment can change the nature of the ambiguity faced by subjects. Specifically, if subjects view the decisions as related, then they would view the experiment as a grand lottery. The different decisions could be interpreted as revealing subjects’ preferences of grand lotteries with different ambiguity properties. This would greatly complicate the analysis.

To see this, consider the example of a typical three–color Ellsberg urn experiment with 30 balls. Subjects know that there are 10 green balls, and that the remaining 20 balls can be either blue or yellow. Typically, subjects are now asked to make a decision between betting on green (G) and betting on, say, blue (B), where the “act” G means that a subject receives $1 if a green ball is drawn and 0 otherwise (see Table 2). Then subjects are asked to make a second decision, namely betting on “not green” (¬G) or on “not blue” (¬B). The standard choice of ambiguity averse subjects would be (G, ¬G), i.e. betting on “green” in the first decision and on “not green” in the second.

Indeed, viewed separately, betting on green is the only unambiguous option for the first decision, and betting on “not green” is the only unambiguous option for the second decision. However, viewed together, betting for and against the same color, no matter which of the colors, produces an unambiguous expected value. As an extreme case, suppose subjects believe that only one ball is drawn, that the color of this ball is relevant for both bets, and that both bets are being paid out. Then, the combination of bets (G, ¬G) yields a guaranteed fixed payment of $1. But so does the bet (B, ¬B).

A well designed ambiguity experiment would, of course, avoid this mistake. So, let us suppose that two independent draws are made from the urn, and that one of the two decisions is selected randomly (e.g. by a fair coin) for payment. However, this procedure still does not solve the problem. In Oechssler and Roomets (2014), we show formally that a decision maker who follows the MaxMin Expected Utility (MEU) approach by Gilboa and Schmeidler (1989), and who considers the two draws from the urn as independent, should be indifferent between the combination of bets (G, ¬G) and (B, ¬B). To avoid these problems, we ask each subject to make only one decision.

Further, this one decision is between a bet on green and one of the two ambiguous colors, where we randomly assign subjects to choosing either between green and blue or between green and yellow. Some experiments (see, e.g. Halevy, 2007) have allowed subjects to choose the color they would like to bet on. This is a very reasonable design choice if one

---

5 Likewise, the distribution of colors of M&M’s can be estimated using pretests and/or by obtaining information from the producer (Chow and Sarin, 2002, did both).

6 Although a bingo blower as employed by Hey et al. (2010) is an ingenious and transparent way of inducing ambiguity, it still induces strategic ambiguity since someone (usually the experimenter) must have filled the blower with differently colored balls and knows the true distribution.

7 This mechanism has been called pay-one decision–randomly (POR), random decision mechanism, or random lottery incentive mechanism. See Cox et al. (2015) and Azrieli et al. (2012) for discussions of the incentives properties of this mechanism.

8 This is also shown by Azrieli et al. (2012) who show that POR is not incentive compatible in the context of ambiguity if the order reversal axiom (see, e.g. Seo, 2009) holds. In our thought experiment, this axiom essentially requires that decision makers do not care whether the balls are drawn first or the coin is tossed first.
wants to alleviate subjects concern about the experimenter trying to trick them. However, as argued already by Raiffa (1961) and more recently by Kuzmics (2013), it allows subject to create their own grand lottery by “flipping a coin in their head” to choose between the two ambiguous colors to create an unambiguous expected value. In a worst-case scenario, subjects could strategically flip coins to eliminate exposure to ambiguity entirely. Since one of our goals is to observe possible differences between subjects exposed to mechanical ambiguity and subjects exposed to strategic ambiguity, it is of paramount importance that we actually expose them to at least one unavoidably ambiguous expected value.

An important drawback of using a single decision between two of the three colors is that strictly speaking we cannot classify individual subjects as ambiguity averse without making an informational symmetry assumption. In other words, we have to assume that the set of priors for blue and yellow is symmetric. Because of this, we focus on aggregate data. Further, we need to make the additional assumption that these priors are distributed similarly regardless of which two colors the subject could bet on, and so we ask about subjects’ priors in a questionnaire at the end of the experiment.

3.3. Dealing with indifference

Anyone who employs the principle of insufficient reason in a typical Ellsberg experiment, as described above, would be indifferent between betting on any color, because such a person would believe that there is an equal expected number of blue and yellow balls and hence 10 of each color in expectation. The phenomenon of ambiguity would not be very relevant if it were restricted to such cases as it may be quite forgivable, even for a Bayesian, to use the fact that the number of green balls is known as a tie-breaker. For this reason, a number of papers have explored whether subjects actually have a strict aversion to ambiguity (see, e.g. Cohen et al., 1985; Curley et al., 1986; Einhorn and Hogarth, 1986; Curley and Yates, 1989; Charness et al., 2013). We chose to address this problem by having 9 of 30 (or 11 of 30, depending on the treatment) green balls in an urn. This way, a decision maker who employs the principle of insufficient reason will not be indifferent.

3.4. Experimental procedures

We used a three–color Ellsberg urn with 30 marbles that were either green, yellow, or blue. The number of green marbles was shown to be 9 (or 11 depending on the treatment). The number of yellow and blue marbles was determined by a newly designed machine. To isolate mechanical ambiguity, we needed to make sure that no one, not even the experimenter could know beforehand the distribution of marbles in the urn. We achieved this by constructing the aforementioned irregular Galton box for bouncy balls, which consists of a vertical board with interleaved rows of nails. To make it transparent to subjects that the experimenters could not have tested the machine before the experiment, the nails were hammered into the board by the subjects themselves at the beginning of the experiment. The timing was as follows:

1. Without knowing anything about the experiment, subjects were asked to volunteer for hammering nails into a plywood board (see Fig. 1, left panel). Black rectangles on the board indicated the areas where a nail had to go but left enough choice for subjects to find a location for their nails. No subject hammered more than two nails.
2. The rest of the machine was brought into the room and the board with the nails was affixed as the back wall of the machine (see Figure Fig. 1, right panel).
3. Subjects were then told that at the end of the experiment one of them would drop 21 (or 19, depending on the treatment) bouncy balls into an opening at the top of the machine. Balls would bounce left and right as they hit the nails. Eventually, they would be collected into different bags at the bottom. There were an equal number of white and black bags with alternating order. The total number of balls in the white bags would then determine the number of yellow marbles that would be put into the urn, while balls in black bags would determine the number of blue marbles.
4. Before the balls were dropped into the machine, subjects had to make one and only one decision. Some were asked to bet either on green (with a known number of marbles) or on yellow, the others were asked to bet either on green or blue. If their color was drawn, a subject would win 6 euro.
5. After the decision sheets were collected, subjects were asked to fill in a questionnaire that asked for their gender and for their expectation with respect to the number of green (as a check whether they understood the instructions), yellow, and blue marbles. They were also asked to explain in a few words how they came to their decision. Finally, there was a question whether they believed that the experimenters could have manipulated the distribution of balls. The questionnaire was unpaid.

---

5 We chose instead to address this issue of strategic ambiguity with our machine.
6 If asymmetric priors were allowed, any behavior would be compatible with expected utility.
7 Of course, all experiments that elicit valuations also account for indifference.
8 From here on we use “marbles” to refer to objects in the Ellsberg urn, while “balls” will denote bouncy rubber super-balls used in our Galton box.
9 Subjects were allowed to verify the bags’ contents after the experiment and some did.
10 For each of our three sessions a new but identical board was used.
11 For a short video of the machine in action see http://www.uni-heidelberg.de/md/awi/professuren/with2/video.mp4
12 In the control treatment C9 (explained below) we asked whether the distribution of marbles could have been manipulated.
6. Finally, one subject would draw a marble from the urn containing 30 marbles.

The experiment was run in 2012 as a classroom experiment with paper and pencil. We chose a classroom experiment since, for reasons explained above, we wanted to run a one-shot experiment where subjects would just make a single decision. To keep opportunity costs of subjects low, we chose to run the experiment at the end of first- or second-year undergraduate economics courses at the Universities of Heidelberg and Mannheim. Since subjects were already present, their opportunity costs of participating were lower than usual. All subjects received a show-up fee of 4 euros. The experiment was a surprise to students. Participation was voluntary but no subject refused to participate. The instructions (see Supplementary material) were distributed on paper. The experiments lasted about 30 min. By design, the earnings were either 10 or 4 euros. The average payment was 6.47 euros.

Our choice to use large classrooms forced us to make a couple concessions. First, we were not able to maintain the level of anonymity we would have had in a more standard laboratory setting. This is true both during the decision-making process, and during the payment process. This allows for the fear of negative evaluation discussed in Trautmann et al. (2008). However, this concern would exist in all of our treatments. A larger concession was that we could only afford to use one class per treatment. Due to the small number of large economics lectures at Heidelberg we were even forced to go to Mannheim to run the C9 treatment. We therefore cannot technically separate the class effect from the treatment effect. We would argue, however, that with regards to the phenomena of interest, ambiguity attitudes, there should be little difference after a year of university education or a 0 min drive. That said, this opens a fair criticism and should be considered when weighing our conclusions.

The experimental treatments were as follows (“M” stands for “mechanical”, “C” for “control”):

- Treatment M9: The known number of green marbles was 9. The machine was used in the classroom to determine the remaining marbles (121 subjects participated).
- Treatment M11: The known number of green marbles was 11. The machine was used in the classroom to determine the remaining marbles (63 subjects).
- Treatment C9: The known number of green marbles was 9. The machine was used privately beforehand without subjects seeing it. Subjects were simply told that “it is unknown how many of the marbles are yellow and how many are blue.” (53 subjects).

The control treatment C9 is the classical 3-color Ellsberg urn experiment as it has been performed many times, with the exception that the percentage of known (green) marbles was not 1/3 but rather 9/30. We chose this design to make sure that subjective expected utility maximizers who have a belief (e.g. because of the principle of insufficient reason) that the two unknown colors blue and yellow are equally likely would have a strict incentive to bet on the unknown color. This way we can classify all those subjects who choose the known color green as ambiguity averse. Treatment M9 is the same as C9 except that a possible strategic aspect of ambiguity, namely that subjects believe that they play a game against the experimenter, should be eliminated through the use of the machine. Finally, treatment M11 was designed to classify subjects as ambiguity

\[17\] Students in Heidelberg and Mannheim (located about 25 km from each other) are likely to be similar in many respects. We chose to add subjects from Mannheim for the control treatment since we had already approached all first- and second-year economics students in Heidelberg. Students from higher years or non-economics students would likely have been more different from our original sample.
Table 3
Percentage of subjects choosing the different bets.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>C9</th>
<th>M9</th>
<th>M11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bet on green</td>
<td>37.7%</td>
<td>45.5%</td>
<td>76.2%</td>
</tr>
<tr>
<td>No. of observations</td>
<td>53</td>
<td>121</td>
<td>63</td>
</tr>
</tbody>
</table>

Notes: We interpret a bet on green as indicating ambiguity aversion in the C9 and M9 treatments. For the M11 treatment, a bet on green indicates ambiguity aversion or ambiguity neutrality.

Table 4
The reasons subjects gave for their decisions.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Reasoning</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td>M11</td>
<td>Principle of insufficient reason</td>
<td>32 1</td>
</tr>
<tr>
<td></td>
<td>Green is safer</td>
<td>8 0</td>
</tr>
<tr>
<td></td>
<td>Blue more likely</td>
<td>0 1</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>8 13</td>
</tr>
<tr>
<td>M9</td>
<td>Principle of insufficient reason</td>
<td>10 30</td>
</tr>
<tr>
<td></td>
<td>Green is safer</td>
<td>25 0</td>
</tr>
<tr>
<td></td>
<td>Blue more likely</td>
<td>0 2</td>
</tr>
<tr>
<td></td>
<td>Yellow more likely</td>
<td>0 6</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>20 28</td>
</tr>
<tr>
<td>C9</td>
<td>Principle of insufficient reason</td>
<td>0 25</td>
</tr>
<tr>
<td></td>
<td>Green is safer</td>
<td>15 0</td>
</tr>
<tr>
<td></td>
<td>Blue more likely</td>
<td>0 7</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>5 7</td>
</tr>
</tbody>
</table>

Notes: The table is based on classifying answers to the question "Why have you decided the way you have?". Category "other" consists mainly of subjects stating that they decided randomly or "by intuition" but also 4 cases where the two coders differed.

loving. All expected utility maximizers who have a belief that the two unknown colors blue and yellow are equally likely would have a strict incentive to bet on the known color green.

4. Experimental results

Before coming to our main results, we check whether the colors of the marbles the subjects could choose from had any effects. We can test for this by comparing groups of subjects who had the choice between green and blue to those who had the choice between green and yellow. Table 7 in the Supplementary material shows the distribution of subjects choice depending on their options (green vs. blue or green vs. yellow). Fisher’s exact tests show that there is no significant difference according to subjects’ options (p-value = 0.47 for M9 and p = 1 for C9). Thus, from now on we pool these data.

Table 3 presents our main results, the percentage of subjects choosing to bet on the color "green". Recall that in treatments C9 and M9 we interpret this percentage as a lower bound on the percentage of ambiguity averse subjects. In both of these treatments we find that the percentage of ambiguity averse subjects is around 40%. The difference between the treatments is not significant according to a χ²-test (two-sided p = 0.34). Thus, interestingly, mistrust of the experimenter seems less of a concern than typically thought by experimentalists and theorists. When strategic ambiguity is more salient (as in a game), this conclusion might change.

Result: The frequency of ambiguity averse subjects is about 40%. There does not seem to be an increase in ambiguity aversion due to strategic interaction with the experimenter.

It is tempting to interpret the percentage of subjects in treatment M11 who did not choose to bet on green as the percentage of ambiguity loving subjects, which would be quite high at 23.8%. However, the questionnaire data may help to put this number into perspective (see Table 4). Of the 15 subjects in M11 who chose to bet on blue, one explicitly says in the questionnaire that he thought that a blue ball was more likely to be drawn. Thus, his behavior would also be compatible with subjective expected utility maximization if his prior of blue is high. All but one of the remaining 14 subjects report that

---

18 We hope to identify subjective expected utility maximizers with an asymmetric prior (e.g. those that believe that there are many more blue marbles in the urn than yellow ones) with the help of our questionnaire. In fact, there were few cases like this in the data.

19 Due to an error by the experimenters this comparison can only be done for treatments M9 and C9. In M11 always the same option (blue vs. green) was used.

20 It is a lower bound since there may be subjects who are mildly ambiguity averse but not so much that this compensates for the fact that the probability of green is less than 1/3.

21 The questionnaire data were categorized by two independent coders who had no previous knowledge or involvement with the experiment. For results of the questionnaire see Table 6 in the Supplementary material.
they made their decisions either randomly or based on “intuition”. Not a single subject expressed a preference for ambiguity (although this may be asking for too much).

In general, Table 4 shows a remarkable consistency of subjects’ decisions and their reasonings. A large majority of ambiguity averse subjects (those choosing green in M9 and C9) cite the belief that “green is safer”. In all treatments, between 33% and 62% of subjects cited the belief that the unknown colors are equally frequent or likely, which is compatible with the principle of insufficient reason, although few if any mentioned the principle. Most of these subjects chose their bets accordingly (i.e. green in M11, and not green in M9 and C9).

To study this relationship further, we present, in column (1) of Table 5, a logit regression for the probability of choosing to bet on green. The variable “principle of insufficient reason” is the product of a corresponding reasoning dummy (that equals one if a subject’s reasoning was classified as indicative of the principle of insufficient reason) multiplied by either 1 or −1, depending on the treatment (1 for M9 and C9, and −1 for M11). Furthermore, we include a gender dummy, a dummy for subjects’ belief (as indicated on the questionnaire) that the experiment could not be manipulated, a dummy for the option a subject had (blue vs. green or yellow vs. green) and treatment dummies.

The reasoning variable is highly significant and has the expected sign. Subjects who have beliefs consistent with the principle of insufficient reason are 40% more likely to choose the option predicted by that principle. Neither gender, nor the belief with respect to the possibility of manipulation, nor the color option subjects were facing, seem to matter.

Regression (2) in Table 5 reports results from a similar regression where we drop the reasoning variable. Here only the treatment dummy for M11 is significant. Looking at the two regressions side by side, it seems that the significance of the M11 dummy in regression (2) can be attributed mainly to subjects who use the principle of insufficient reason that supports the ambiguous choice in M9 and C9, but support the risky choice in M11.

4.1. Discussion of results and contributions

In discussing the impact of our results it is important to keep in mind the design goals of our experiment. We attempted as much as possible to create a design that transparently removed our ability to influence the distribution of the contents of the urn in any predictable way. We used this mechanical procedure to measure both ambiguity aversion and ambiguity seeking. We further use a non-mechanical, “standard” procedure to measure ambiguity aversion. Because our results are consistent with existing literature it may seem at first pass that they do not add much to the discussion. However, because of the factors we control for, our results can speak to the psychological source of ambiguity in ways other papers have not. The lack of statistical differences between the M9 and C9 treatments suggests that the types of psychological motives present when confronted with strategic ambiguity (being tricked or duped, e.g.) are not driving aversion in standard designs.

Furthermore, while we may not be the first to attempt to remove one’s ability to influence the distribution, we do believe that we have done it most completely. Our results can be viewed then as currently the least strategically influenced measurement. This has value in itself, but also allows others who may be interested in pursuing similar questions either a design example to follow, or at least exposure to the types of issues one has to address when developing alternative designs.

5. Conclusion

Our results suggest that a significant portion of society has an aversion to mechanical ambiguity. Further, we do not find a significant difference between the proportion who are averse to mechanical ambiguity and the proportion averse to strategic ambiguity. These results suggest that the psychological underpinnings of ambiguity aversion may be primarily non-strategic in nature. This conclusion is bolstered by our survey data that shows that similar choices in strategic and mechanical treatments were supported by similar reasoning processes. In particular, ambiguity neutral choices were often

---

22 We could not include the dummy variable “green is safer” (which equals one if a subject’s reasoning focussed on the safety of the green bet) because as Table 4 shows, in 100% of cases where subjects stated that green was safer, they did choose green.
backed by the principle of insufficient reason, while ambiguity averse decisions seemed to be supported by the relative "safety" of the risky choice.

Further, our three design features (the machine, a single decision, and strict preferences) ensure that what we observe is as close to mechanical ambiguity as has been seen in the lab. This is important not only for the applicability of our results, but, to an extent, the broader ambiguity literature as well. Our observed levels of ambiguity aversion are in line with much of the existing literature, though on the lower end. Thus, in the aggregate, our design features may have produced more conservative results when compared with a standard design. Overall, however, the differences are not too large: even when isolating mechanical ambiguity and controlling for ambiguity hedging and indifference, we find that around 40% of subjects seem to be ambiguity averse.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jebo.2015.08.004.

References


