

**EXPERIMENTAL EVALUATION  
OF RANDOM INCENTIVE SYSTEM  
UNDER AMBIGUITY**

Tomohito Aoyama  
Nobuyuki Hanaki

Secondly Revised September 2025

Revised May 2025

March 2024

The Institute of Social and Economic Research  
Osaka University  
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

# Experimental Evaluation of Random Incentive System under Ambiguity \*

Tomohito Aoyama<sup>†</sup>

Nobuyuki Hanaki<sup>‡</sup>

## Abstract

The random incentive system (RIS) is a standard incentive scheme used to elicit preferences in economic experiments. However, it has been shown that RIS may distort observed preferences. We examine the performance of RIS under ambiguity with two sets of experiments, our own and another replicating the main treatments of Baillon et al. (2022a). Contrary to Baillon et al. (2022a), who report a significantly lower proportion of participants revealing strict ambiguity aversion in the treatment with RIS than the one without, we do not find such evidence either in our own or in replication of Baillon et al. (2022a).

KEYWORDS: Random incentive system, Incentive compatibility, Ambiguity, PRINCE

JEL CLASSIFICATION: C91, D81

---

\*This is a major revision of a paper circulated under the same title. The current version mainly reports the results of the newly conducted laboratory experiments, including those that replicate the main treatments of Baillon et al. (2022a). The description and results of earlier experiments (reported in the previous version) are all presented in the online supplementary material of the paper. We gratefully acknowledge financial support from the Joint Usage/Research Center at the Institute of Social and Economic Research (ISER), the University of Osaka, and the Japan Society for the Promotion of Science (18K19954, 20H05631, 20K22088, 23H00055, 25H00388). We also thank the experimental economics laboratory at the Research Institute for Socionetwork Strategies at Kansai University, the BKC Research Organization of Social Sciences at Ritsumeikan University (especially Ai Takeuchi), and the Division of Policy and Planning Sciences Laboratory at Tsukuba University (especially Eizo Akiyama) for their support and hospitality in conducting the experiments. The support by Satsuki Yamada and Yuta Shimodaira in conducting experiments is also gratefully acknowledged. Yuhao Fu has provided an excellent research assistance. The experiments reported in this paper were approved by the IRB of ISER, the University of Osaka (dated 2021/11/08, No. 20220601, 20240905) and are pre-registered at <https://aspredicted.org/6wtx-7y8p.pdf> for the main experiment and <https://aspredicted.org/skrz-tbr6.pdf> for the replication of the experiment in Baillon et al. (2022a).

<sup>†</sup>Non-academic institution. E-mail: [tomo.aoyama.jp@gmail.com](mailto:tomo.aoyama.jp@gmail.com)

<sup>‡</sup>Institute of Social and Economic Research, the University of Osaka, Japan, and the University of Limassol, Cyprus.  
Email: [nobuyuki.hanaki@iser.osaka-u.ac.jp](mailto:nobuyuki.hanaki@iser.osaka-u.ac.jp)

# 1. Introduction

Since Ellsberg (1961), many studies have sought to identify individuals' ambiguity attitudes by experimentally implementing the Ellsberg paradox (Trautmann and van de Kuilen, 2015), which involves tasks in which participants choose between ambiguous and risky bets. To incentivize these tasks, rewards are often determined by the choice made in a randomly selected task. This incentive scheme, known as the Random Incentive System (RIS), is widely used under the assumption that it is incentive compatible—that is, that observed choices reflect underlying preferences.<sup>1</sup>

However, theoretical work has shown that RIS may fail to be incentive compatible when participants exhibit nontrivial attitudes toward ambiguity, as it allows them to hedge ambiguity through randomness (Raiffa 1961; Oechssler and Roomets 2014; Bade 2015; Kuzmics 2017). While Baillon et al. (2022b) and Monet and Vergopoulos (2024) demonstrate theoretically that incentive compatibility may be restored if randomization occurs before decisions are made—as in the PRINCE method proposed by Johnson et al. (2021)—Baillon et al. (2022a) provide experimental evidence to the contrary. In their laboratory study using student samples, they report that the proportion of participants revealing strict ambiguity aversion under RIS was roughly half that observed without RIS, regardless of whether randomization occurred before or after the decision.<sup>2</sup> Specifically, Baillon et al. (2022a) report that strict ambiguity aversion occurred in 50% (42 of 84) of participants without RIS, 28.7% (25 of 87) with RIS when randomization occurred before the choice, and 25.3% (22 of 87) with RIS when randomization occurred after the choice. If this result is robust, it challenges the use of RIS to experimentally identify ambiguity attitudes.<sup>3</sup>

The present study reexamines the incentive compatibility of RIS under ambiguity by comparing choices in treatments with and without RIS in a design similar, though not identical, to Baillon et al.

---

<sup>1</sup>Other incentive schemes have also been employed in experimental economics. For instance, participants are sometimes paid for all choice situations. Azrieli et al. (2018) show that this approach requires a stronger preference condition to achieve incentive compatibility than RIS.

<sup>2</sup>According to Johnson et al. (2021), PRINCE methods employ four principles: (1) PRiority (the real choice situation, RCS, is determined at the start of the experiment before any decision is made), (2) INstructions to experimenter (participants provide explicit instructions for what the experimenter should do in the RCS), (3) Concreteness (participants receive a tangible description of the RCS, such as a sealed envelope), and (4) Entirety (participants receive a full and unambiguous description of the RCS).

<sup>3</sup>RIS incentive compatibility under risk has also been questioned. Holt (1986) and Karni and Safra (1987) predicted its failure if the independence axiom of expected utility is violated. Subsequent experiments tested this prediction (Cubitt et al., 1998; Cox et al., 2014, 2015; Harrison and Swarthout, 2014; Brown and Healy, 2018; Freeman et al., 2019); except for Cubitt et al. (1998), all found supporting evidence. Notably, Brown and Healy (2018) show that RIS distorts revealed risk preferences when risky choices are presented in an ordered multiple-choice list on a single page but not when presented one per page in random order.

(2022a). Specifically, in both our online and main laboratory experiments, we test monotonicity—a preference condition identified by Azrieli et al. (2018) as sufficient for RIS incentive compatibility.<sup>4</sup> Additionally, we replicate the three main treatments of Baillon et al. (2022a) with a sample size sufficient to achieve 90% statistical power to detect their reported effect size at the 5% significance level.

Contrary to Baillon et al. (2022a), we find no significant differences in the proportion of subjects revealing strict ambiguity aversion between RIS and non-RIS treatments, either in our own setting (conducted both online and in the laboratory) or in our replication of their experiment. In our main laboratory study with over 660 participants (more than 110 per treatment), the proportion revealing strict ambiguity aversion was 44.2% (50 of 113) in the treatment without RIS and 49.1% (57 of 116) in the treatment with RIS in our own design. In our replication of Baillon et al. (2022a), these proportions were 35.5% (39 of 110) without RIS, 45.5% (51 of 112) with RIS where randomization occurred before the choice, and 40.5% (45 of 111) with RIS where randomization occurred after the choice.

We conjecture that the divergence between our findings and those of Baillon et al. (2022a) stems from differences in subject pools, particularly in their tendency to bracket choices. As noted by Baillon et al. (2022a), RIS fails to be incentive compatible when participants bracket broadly—evaluating choices jointly across tasks—but remains compatible when participants bracket narrowly and evaluate each task independently. Experimental studies on choice bracketing (Tversky and Kahneman, 1981; Rabin and Weizsäcker, 2009; Ellis and Freemand, 2024) suggest that many individuals bracket choices narrowly, though this proportion vary substantially—from 28% in Rabin and Weizsäcker (2009) to about 70% in Ellis and Freemand (2024)—across subject pools and experimental designs. Thus, it is plausible that participants in Baillon et al. (2022a) were more prone to broad bracketing than those in our study. This highlights the importance of understanding a subject pool’s bracketing tendencies when using RIS to elicit ambiguity attitudes.

We also find that results from our online and laboratory experiments, both conducted with student samples, are highly similar despite differences in stake size.<sup>5</sup> For example, in our laboratory experiment, 54.1% (60 of 111) chose the risky bet over the ambiguous bet without RIS, compared to 62.1%–64.7% (72 or 75 of 116) with RIS (depending on bet color). In Session I of Online Experiment I, the proportions were 64.6% (166 of 257) without RIS and 68.6% (107 of 156) with

---

<sup>4</sup>The settings and results of the online experiments are detailed in the online supplementary material.

<sup>5</sup>In the laboratory experiment, stake sizes were doubled (1,000 JPY vs. 500 JPY) and participation fees halved (500 JPY vs. 1,000 JPY) relative to the online experiment.

RIS (see Section I.4.2 of the supplementary material). In Online Experiment II, 58.3% (60 of 103) revealed strict ambiguity aversion without RIS, compared to 58.3% (91 of 156) and 53.7% (65 of 121) in Sessions I and II of Online Experiment I, respectively (see Section II.3 of the supplementary material). These findings suggest that, for simple tasks, conducting experiments online does not inherently bias results.

Although online experiments have become increasingly popular since the COVID-19 pandemic, their comparability with laboratory studies remains debated. Prior work comparing online and laboratory sessions with participants from the same pool reports mixed findings: some detect no significant differences (Hergueux and Jacquemet, 2015; Ozono and Nakama, 2022; Snowberg and Yariv, 2021; Hanaki et al., 2022), while others find notable discrepancies (Prissé and Jorrat, 2021; Schmelz and Ziegelmeyer, 2020). Our results add further evidence supporting the comparability of online and laboratory data for simple ambiguity tasks.

The remainder of the paper is organized as follows: Section 2 describes the experimental design, Section 3 presents the results, and Section 4 offers concluding remarks.

## **2. Experiments**

### **2.1. Main experiment**

In our main experiment, participants are presented with two boxes, Box A and Box B, as depicted in Figure 1. Box A contained 9 green balls and 11 yellow balls. We used 9 green balls instead of 10 for tie-breaking, which prevents ambiguity-neutral participants from choosing Bet G in both questions. This technique was used by Oechssler and Roomets (2015). The balls in Box A were disclosed publicly in the laboratory so that all participants could verify them. Box B contained 20 balls that were either red or black, with the numbers of each color determined by students who participated in another experiment. Although participants were informed that the color composition was determined in this way, they were not informed of the exact number of each color.

Participants were presented with the following three bets.

Bet G: If a ball taken out of Box A is green, 1000 JPY is paid out,

Bet R: If a ball taken out of Box B is red, 1000 JPY is paid out,

Bet B: If a ball taken out of Box B is black, 1000 JPY is paid out.

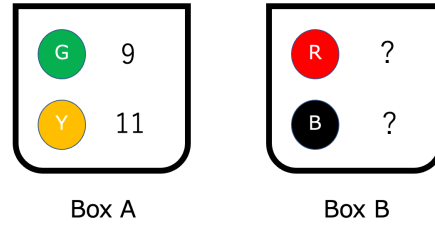


Fig. 1. The two boxes presented to participants.

Otherwise, participants received nothing. In addition, all participants received a 500 JPY participation fee. Then, they were asked to answer the following two questions:

Question Red: Which do you prefer, Bet G or Bet R?

Question Black: Which do you prefer, Bet G or Bet B?

Participants received instructions about both questions before making their choices so they could view the entire experiment as a single decision problem. They made their choices on a separate answer sheet, in which each question was displayed on one side of the paper or the other.

To ensure transparency, at the beginning of the experiment, we selected one participant in each session (who was the last one to arrive) as an implementer who received a fixed payment of 1000 JPY, including a 500 JPY participation fee. The implementer was not informed of the nature of the problem that other participants faced and waited in another room while all the other participants received instructions and made their own decisions. Once all the participants in the laboratory made their decisions and their answer sheets were collected by the experimenter, the implementer was invited back to the laboratory and drew a ball from each box in front of all the other participants (without seeing the content of the ball by wearing an eye mask) to determine the payment.

There are two main treatments: ONE and TWO. In treatment ONE, although participants answer both questions, they are informed of their payment-relevant question (which we call the real choice situation, or RCS) when making their decisions. Thus, half of the participants in the laboratory receive an instruction stating that they will be paid according to their choice in question Red, and the other half receive another instruction stating that they will be paid according to their choice in question Black. This fact is made clear to all the participants when the experimenter reads the instructions aloud and asks participants to check the RCS written on their instructions.

In treatment TWO, participants answer both questions but are paid based on their choice of one of the two questions. The RCS is determined by the implementer in a separate room at the

beginning of the experiment. After moving to that room, the implementer throws a six-sided die (with three faces in red and three faces in black) once to determine the RCS for all the participants in the session. The sealed envelopes containing the selected RCS are brought to the laboratory by an assistant and are distributed to participants by the experimenter before they make their choices. Participants are clearly informed about this procedure in their instructions. Once all the participants have received an envelope, they make their choices without knowing which question is the RCS.<sup>6</sup> Participants are asked to open the envelope only after all participants have made their choices and the experimenter has collected all the answer sheets.

Note that a strictly ambiguity averse participant, who would choose Bet G in the RCS in treatment ONE, can benefit by choosing Bet R and Bet B in questions R and B in treatment TWO. This is because as shown in Appendix A, by doing so, they can construct a compound lottery that gives them 1000 JPY with 50% probability. This compound lottery is better than choosing Bet G in both questions that gives them 1000 JPY with 45% probability. We will, therefore, compare the fraction of participants who have chosen Bet G in question R between treatment TWO and those in treatment ONE whose RCS was R, as well as in question B between treatment TWO and those in treatment ONE whose RCS was B.

In addition, we have a treatment called SINGLE, which is similar to the treatment Single in Baillon et al. (2022a), which is described in the next subsection. In this treatment, participants first choose which question, Red or Black, to answer, and only make a choice on the question they have chosen. This treatment is added to compare the proportion of participants who reveal strict ambiguity aversion ( $SAA^r$ ) by choosing Bet G in their chosen problem in the treatment without RIS (SINGLE) or by choosing Bet G in both problems in treatment with RIS (TWO), as in Baillon et al. (2022a).

Experiments are not computerized. Participants submit their choices on sheets of paper. Randomization is done by a participant who is selected as an implementer with a six-sided die (in determining RCS in treatment TWO) and physical boxes that contain twenty balls. See Appendix B for an English translations of the instructions.

## 2.2. Replication of Baillon et al. (2022a)

We have also replicated the three main treatments of Baillon et al. (2022a): Single, Before, and After. In all the treatments, participants face the two choice problems shown in Figure 2. There are

---

<sup>6</sup>Thus, we follow the priority and concreteness principle of the PRINCE method (Johnson et al., 2021).

two bags, Bag A and Bag B, each of which contains two chips. Bag A contains one red chip and one blue chip. The chips in bag A are disclosed publicly in the laboratory so that all participants can verify them. The color composition of Bag B is not disclosed to participants. In the choice problem RED (BLUE), participants choose between two options. If they select option a, they obtain 1000 JPY if a red (blue) chip is drawn from Bag A. If they select option b, they obtain 1020 JPY if a red (blue) chip is drawn from Bag B. Otherwise, they receive nothing. In addition, all participants receive a 500 JPY participation fee.<sup>7</sup>

To ensure transparency, at the beginning of the experiment, we selected one participant in each session (who was the last one to arrive) as an implementer who received a fixed payment of 1000 JPY, including a 500 JPY participation fee. The implementer is not informed of the nature of the problem that other participants face and waits in another room while all the other participants receive instructions and make their decisions.<sup>8</sup> Once all the participants in the laboratory have made a decision and their answer sheets have been collected by the experimenter, the implementer is invited back to the laboratory and draws a chip from each bag in front of all the other participants (without seeing the contents of the bag by wearing an eye mask) to determine the payment.

In the treatment Single, participants first choose which problem, RED or BLUE, to answer, and only make a choice for the problem they have chosen.

In treatment Before, participants make their choices for both problems, but they are paid according to their choice for one of the problems (i.e., in an RCS). The RCS is selected randomly before they make their choices: at the beginning of the experiment, the implementer in the separate room rolls a six-sided die (with three red faces and three blue faces) for each participant.<sup>9</sup> Sealed envelopes that contain the resulting RCSs are distributed to the participants in the laboratory by the experimenter. Once all participants receive an envelope, they make their choices without knowing their RCS. Participants are asked to open the envelope only after all the participants have made their choices and the experimenter has collected all the answer sheets.

---

<sup>7</sup>In Baillon et al. (2022a), participants received 5 euros as a participation fee, and either 0, 10.0, or 10.2 euros in reward. Although the stake size of Baillon et al. (2022a) is large than our experiment based on the exchange rate when the experiment is conducted (around 1 euro = 160 JPY), it is comparable or even higher under our experiment based on the purchasing power parity (around 1 euro = 85 JPYs based on the consumer price index), according to the Institute for International Monetary Affair (<https://www.iima.or.jp/en/research/ppp.html> accessed on July 30, 2025).

<sup>8</sup>Before moving to the other room, the implementer selects one out of six bags without seeing its contents. That bag is used as Bag B for the session.

<sup>9</sup>This is different from our treatment Two, in which the implementer throws a six-sided die only once for all participants in the session; thus, everyone in the session faces the same RCS. In the treatment Before of Baillon et al. (2022a), because the implementer throws a six-sided die once for each participant, the RCS may differ across participants in the session.

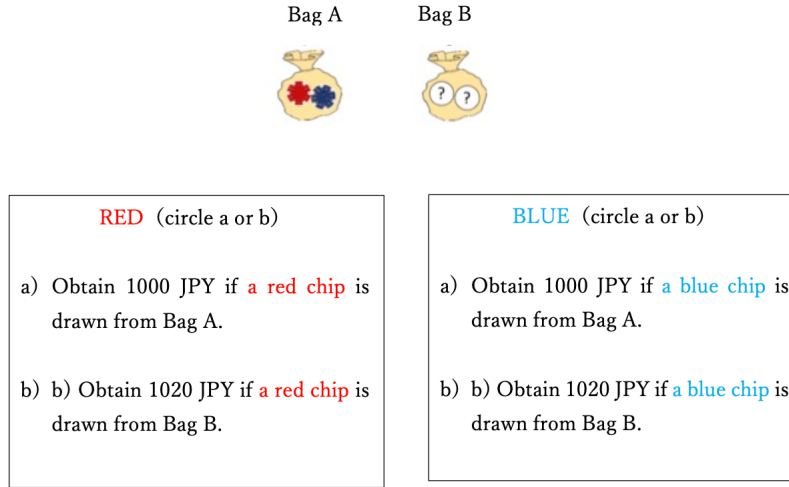


Fig. 2. Two choice problems in our replication of Baillon et al. (2022a).

In treatment After, participants make their choices in both problems, but they are paid according to their choice in one of the problems. The RCS is selected randomly at the end of the experiment. The implementer, after drawing the chips from the bag and announcing their colors, rolls a six-sided die (three red and three blue faces) in front of each participant once to determine their RCSs.

Experiments are not computerized. Participants submit their choices on sheets of paper. Randomization is done by a participant who is selected as an implementer with a six-sided die (to determine the RCS in treatments Before and After) and physical bags that contain two chips.

We are interested in comparing the fraction of participants who reveal strict ambiguity aversion ( $SAA^r$ ) by choosing option a for their chosen problem in treatment Single (the one without RIS) or choosing option a for both problems in treatments Before and After (those with RIS).

### 3. Results of the Experiment

The experiments were conducted between November 2024 and April 2025 in the experimental laboratory at Osaka University, Kansai University, Ritsumeikan University, and Tsukuba University. Participants are recruited from the participant database managed by ORSEE (Greiner, 2015) in Osaka and Kansai and from the students populations via posting on an online bulletin board at Ritsumeikan and Tsukuba. These participants had never taken part in a similar experiment and participated in only one session. Those in Osaka who participated in our earlier experiments (both

Table 1  
Number of participants from each location in six treatments.

|                          | Single | Before | After | SINGLE | ONE   | TWO |
|--------------------------|--------|--------|-------|--------|-------|-----|
| Osaka                    | 42     | 44     | 42    | 44     | 41    | 52  |
| Kansai                   | 31     | 33     | 36    | 31     | 40    | 31  |
| Ritsumeikan              | 23     | 21     | 22    | 23     | 18    | 19  |
| Tsukuba                  | 14     | 14     | 11    | 15     | 12    | 14  |
| Total                    | 110    | 112    | 111   | 113    | 111   | 116 |
| $p$ -values <sup>a</sup> |        | 0.986  |       |        | 0.683 |     |

<sup>a</sup>  $p$ -values based on  $\chi^2$  test.

online and preliminary laboratory experiments reported in the online supplementary material) are excluded from the invitation list. This was the reason why we needed to conduct sessions at various universities to reach the targeted number of observations. All these universities are in Japan.

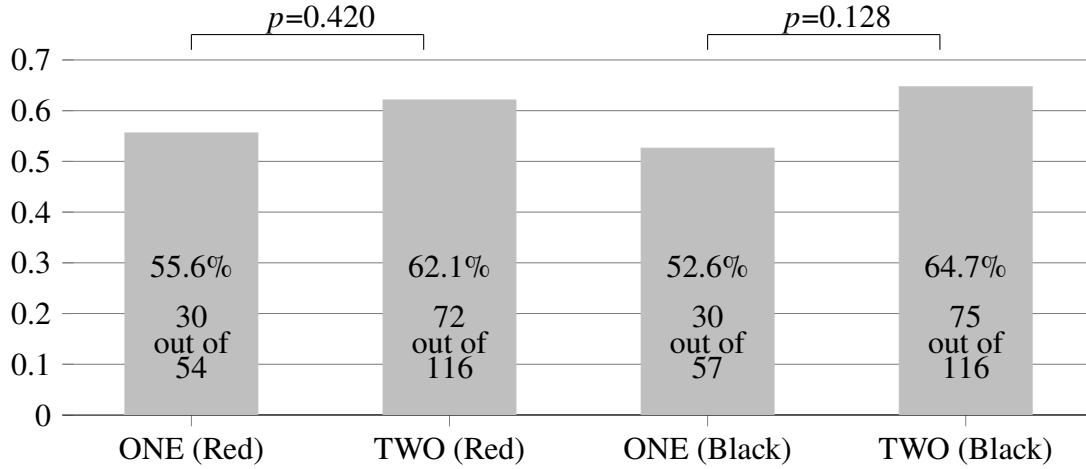
The numbers of participants from each location in each treatment are presented in Table 1. We aimed to collect 109 observations in each treatment to obtain a statistical power of 90% to detect the original effect size reported by Baillon et al. (2022a) at the 5% significance level. As noted in the introduction, Baillon et al. (2022a) report that the fractions of SAA<sup>r</sup> participants are 50%, 28.7%, and 25.3% in Single, Before, and After, respectively.<sup>10</sup> However, the final number of observations varies slightly across treatments due to variations in show-up rates. The fractions of participants from four locations are balanced across three treatments replicating Baillon et al. (2022a) and across three treatments of our own ( $p$ -values are 0.986 for the former and 0.683 for the latter based on the  $\chi^2$  test).

The experiment lasted, on average, 20 minutes, including the instruction and payments. Participants, excluding the implementers who earned 1000 JPY, earned an average of 994.2 JPY, including the 500 JPY participation fee.

### 3.1. Comparison between ONE and TWO

We first compare the fraction of participants who chose Bet G between treatments ONE and TWO. Recall that in treatment ONE, participants were informed of their RCS when making their choices.

<sup>10</sup>In the pre-registration, we stated we would collect 100 observations per treatment, but later we realized that there was an error in the power computation (we based it on the  $\chi^2$  test instead of the proportion test) and 100 observations per treatment was not enough to achieve the power of 90%. Thus, we slightly increased the sample size.



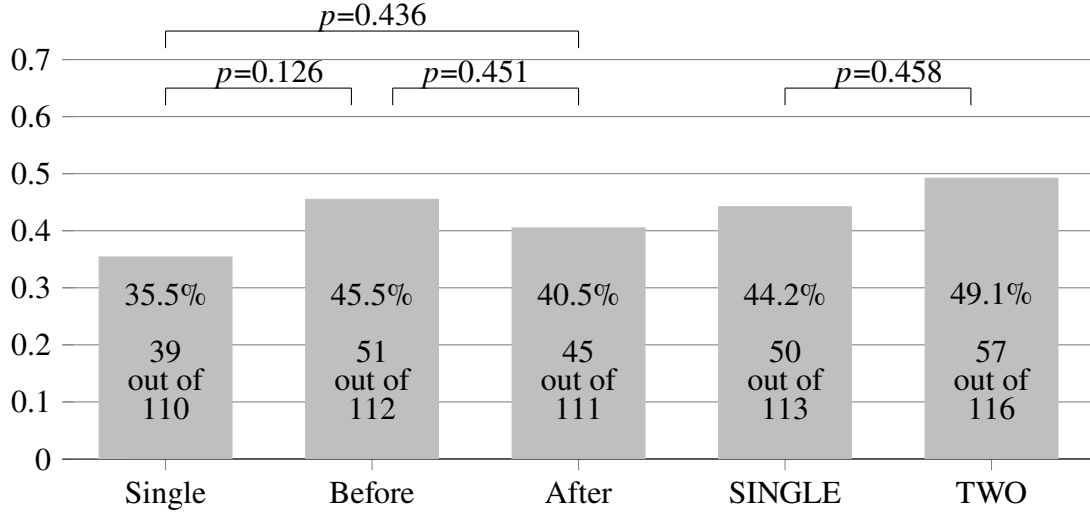
Note:  $p$ -values are based on two sample proportion tests.

Fig. 3. The proportion of participants who have chosen Bet G.

Thus, we only consider choices in their respective RCS in treatment ONE and compare them with the choices observed in the same question in treatment TWO.

Figure 3 shows the results. The fraction of participants whose RCS was question Red (Black) in treatment ONE and chose Bet G was 55.6% (52.6%), while those in treatment TWO and chose Bet G in question Red (Black) was 62.1% (64.7%). These fractions are not statistically significantly different between the two treatments ( $p$ -values are 0.420 for question Red and 0.128 for question Black, two-sample proportion test).<sup>11</sup>

<sup>11</sup> In the last session conducted at Osaka (session No. 2504231330), where treatment TWO was conducted, our assistant inadvertently asked the implementer to throw a six-sided die for each participant, instead of once for all the participants, and prepared envelopes although the instruction for the implementer clearly stated that the die should be thrown only once for everyone. This error was discovered when participants opened their envelopes; that is, after participants had submitted their response and the answer sheet were corrected but before the balls were drawn from the two boxes. Admitting the error and with an agreement with participants, we proceeded to draw balls to determine their payoffs. The data from this session (11 observations) are included in the analyses as the error does not affect the choices. The results do not change even if we drop these observations. If we drop these data, these fractions become 62.9% (66 out of 105) and 65.7% (69 out of 105) for questions Red and Black, respectively, and they are not statistically significantly different from those of ONE ( $p=0.373$  and  $p=0.103$ , respectively for questions Red and Black, proportion tests). In the pre-registration, we stated that we use a one-tailed  $\chi^2$  test with Bonferroni correction for multiple hypothesis testing. We are reporting the results of the proportion test to be consistent with the analyses used in Baillon et al. (2022a). The  $p$ -values (two-tailed  $\chi^2$  test) are 0.420 for question Red and 0.128 for question Black. Note that contrary to our original hypothesis, the fraction of safe choices is smaller in ONE than in TWO for both questions.



Note:  $p$ -values are based on two sample proportion tests.

Fig. 4. The proportion of participants who revealed strict ambiguity aversion in each treatment.

### 3.2. Fraction of $SAA^r$ participants

Figure 4 shows the proportion of  $SAA^r$  participants in the five treatments other than treatment ONE.  $SAA^r$  participants are those who have chosen option a (in case of Single) or Bet G (in case of SINGLE) in the question they have chosen, or those who have chosen option a (in case of Before and After) or Bet G (in case of TWO) in both questions in treatments Before, After, and TWO.

We first discuss the results of the experiment replicating Baillon et al. (2022a). The proportion of  $SAA^r$  participants are 35.5%, 45.5%, and 40.5% in treatments Single, Before, and After, respectively, and they are not significantly different across treatments ( $p$ -values are 0.126 between Single and Before, 0.436 between Single and After, and 0.451 between Before and After, proportion tests). Thus, contrary to Baillon et al. (2022a), we do not observe a significantly lower proportion of  $SAA^r$  participants in the two treatments involving RIS (Before and After) compared to the treatment without it (Single).

Similarly, the proportions of  $SAA^r$  participants in our own treatments, 42.9% in SINGLE and 49.1% in TWO, are not significantly different ( $p=0.333$ , proportion test).<sup>12</sup> Thus, our results suggest that RIS does not distort the choice under ambiguity.

<sup>12</sup>If we drop the 11 observations from the session with an implementation error in treatment TWO noted in footnote 11, the fraction is 49.5%, and it is not significantly different from that of SINGLE ( $p=0.435$ , proportion test). In the pre-registration, we stated that we use a one-tailed  $\chi^2$  test. We are reporting the results of the proportion test to be consistent with the analyses used in Baillon et al. (2022a). The  $p$ -value (two-tailed  $\chi^2$  test) is 0.458. Note that contrary to our original hypothesis, the proportions of  $SAA^r$  participants are smaller in SINGLE than in TWO.

We also note that, while we did not pre-register these comparisons, we do not observe significant differences in the proportions of  $SAA^r$  participants between our original design and that of Baillon et al. (2022a) that shares the same feature; that is, between Before and TWO ( $p=0.586$ , proportion tests) and between Single and SINGLE ( $p=0.180$ , proportion tests).

## 4. Concluding remark

In this study, we investigated the incentive compatibility of the RIS under ambiguity. We conducted two sets of laboratory experiments, as well as set of online experiments reported in the supplementary material. The first experiment, designed by us, tested monotonicity—a preference condition identified by Azrieli et al. (2018) as sufficient for RIS incentive compatibility. The second experiment replicated the main treatments of Baillon et al. (2022a), in which participants made choices over risky and ambiguous bets under different incentive schemes.

Drawing on data from more than 660 participants (over 110 participants per treatment, providing 90% power to detect the 100% effect size reported by Baillon et al. (2022a)), we found no clear evidence that RIS distorts choices under ambiguity in either experimental design. Moreover, our online and laboratory experiments, both conducted with student samples, yielded remarkably similar results despite differences in stake size. Although the use of online experiments has become increasingly common since the COVID-19 pandemic, the comparability of online and laboratory data remains debated. Our findings contribute evidence in favor of their comparability, at least for the relatively simple task examined here.

We conjecture that the divergence between our findings and those of Baillon et al. (2022a) arises from differences in subject pools, particularly in their tendency to bracket choices. As Baillon et al. (2022a) notes, RIS fails to be incentive compatible if subjects bracket broadly but remains compatible when they bracket narrowly. Experimental studies on choice bracketing (Tversky and Kahneman, 1981; Rabin and Weizsäcker, 2009; Ellis and Freemand, 2024) suggest that many individuals bracket choices narrowly, though this proportion vary substantially—from 28% in Rabin and Weizsäcker (2009) to about 70% in Ellis and Freemand (2024)—across subject pools and experimental designs. Thus, it is plausible that participants in Baillon et al. (2022a) were more prone to broad bracketing than those in our study. This highlights the importance of understanding a subject pool’s bracketing tendencies when using RIS to elicit ambiguity attitudes.

Finally, while Ellis and Freemand (2024) show that tendencies to bracket choices correlate across risky portfolio and social allocation tasks, the relationship between bracketing in risky

choices and hedging in ambiguous choices under RIS has not, to our knowledge, been examined experimentally. We view this as a promising direction for future research.

## References

- Anscombe, F. J. and R. J. Aumann (1963) “A Definition of Subjective Probability,” *The Annals of Mathematical Statistics*, 34 (1), 199–205.
- Azrieli, Yaron, Christopher P. Chambers, and Paul J. Healy (2018) “Incentives in Experiments: A Theoretical Analysis,” *Journal of Political Economy*, 126 (4), 1472–1503.
- Bade, Sophie (2015) “Randomization Devices and the Elicitation of Ambiguity-Averse Preferences,” *Journal of Economic Theory*, 159, 221–235.
- Baillon, Aurélien, Yoram Halevy, and Chen Li (2022a) “Randomize at Your Own Risk: On the Observability of Ambiguity Aversion,” *Econometrica*, 90 (3), 1085–1107.
- (2022b) “Experimental Elicitation of Ambiguity Attitude Using the Random Incentive System,” *Experimental Economics*, 25 (3), 1002–1023.
- Brown, Alexander L. and Paul J. Healy (2018) “Separated Decisions,” *European Economic Review*, 101, 20–34.
- Cox, James C., Vjollca Sadiraj, and Ulrich Schmidt (2014) “Asymmetrically Dominated Choice Problems, the Isolation Hypothesis and Random Incentive Mechanisms,” *PLoS ONE*, 9 (3), e90742.
- (2015) “Paradoxes and Mechanisms for Choice under Risk,” *Experimental Economics*, 18 (2), 215–250.
- Cubitt, Robin P., Chris Starmer, and Robert Sugden (1998) “On the Validity of the Random Lottery Incentive System,” *Experimental Economics*, 1 (2), 115–131.
- Ellis, Andrew and David J. Freemand (2024) “Revealing Choice Bracketing,” *American Economic Review*, 114, 2668–2700.
- Ellsberg, Daniel (1961) “Risk, Ambiguity, and the Savage Axioms,” *The Quarterly Journal of Economics*, 75 (4), 643–669.

- Freeman, David J., Yoram Halevy, and Terri Kneeland (2019) “Eliciting Risk Preferences Using Choice Lists,” *Quantitative Economics*, 10 (1), 217–237.
- Greiner, Ben (2015) “Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE,” *Journal of the Economic Science Association*, 1 (1), 114–125.
- Hanaki, Nobuyuki, Takayuki Hoshino, Kohei Kubota, Fabrice Murtin, Masao Ogaki, Fumio Ohtake, and Naoko Okuyama (2022) “Comparing Data Gathered in an Online and a Laboratory Experiment Using the Trustlab Platform,” ISER DP 1168, Institute of Social and Economic Research, the University of Osaka.
- Harrison, Glenn W. and J. Todd Swarthout (2014) “Experimental Payment Protocols and the Bipolar Behaviorist,” *Theory and Decision*, 77 (3), 423–438.
- Hergueux, Jérôme and Nicolas Jacquemet (2015) “Social preferences in the online laboratory: a randomized experiment,” *Experimental Economics*, 18, 251–283.
- Holt, Charles A (1986) “Preference Reversals and the Independence Axiom,” *American Economic Review*, 76 (3), 508–515.
- Johnson, Cathleen, Aurélien Baillon, Han Bleichrodt, Zhihua Li, Dennie Van Dolder, and Peter P. Wakker (2021) “Prince: An Improved Method for Measuring Incentivized Preferences,” *Journal of Risk and Uncertainty*, 62 (1), 1–28.
- Karni, Edi and Zvi Safra (1987) “‘Preference Reversal’ and the Observability of Preferences by Experimental Methods,” *Econometrica*, 55 (3), 675–685.
- Kuzmics, Christoph (2017) “Abraham Wald’s Complete Class Theorem and Knightian Uncertainty,” *Games and Economic Behavior*, 104, 666–673.
- Monet, Benjamin and Vassili Vergopoulos (2024) “Ambiguity, Randomization and the Timing of Resolution of Uncertainty,” *Economic Theory*, 78, 1021–1045.
- Oechssler, Jörg and Alex Roomets (2014) “Unintended Hedging in Ambiguity Experiments,” *Economics Letters*, 122 (2), 243–246.
- (2015) “A Test of Mechanical Ambiguity,” *Journal of Economic Behavior & Organization*, 119, 153–162.

- Ozono, Hiroki and Daisuke Nakama (2022) “Effects of experimental situation on group cooperation and individual performance: comparing laboratory and online experiments,” *Plos One*, 17 (4), e0267251.
- Prissé, Benjamin and Diego Jorrat (2021) “Lack of Control: An experiment,” Technical Report 3932079, SSRN.
- Rabin, Matthew and Georg Weizsäcker (2009) “Narrow Bracketing and Dominated Choices,” *American Economic Review*, 99, 1508–1543.
- Raiffa, Howard (1961) “Risk, Ambiguity, and the Savage Axioms: Comment,” *The Quarterly Journal of Economics*, 75 (4), 690–694.
- Schmelz, Katrin and Anthony Ziegelmeyer (2020) “Reactions to (the absence of) control and workplace arrangements: experimental evidence from the internet and the laboratory,” *Experimental Economics*, 23, 933–960.
- Snowberg, Erik and Leeat Yariv (2021) “Testing the Waters: Behavior across participant pools,” *American Economic Review*, 111 (2), 687–719.
- Trautmann, Stefan T. and Gijs van de Kuilen (2015) “Ambiguity Attitudes,” in Keren, Gideon and George Wu eds. *The Wiley Blackwell Handbook of Judgment and Decision Making*, 89–116, Chichester, UK: John Wiley & Sons, Ltd.
- Tversky, Amos and Daniel Kahneman (1981) “The Framing of Decisions and the Psychology of Choice,” *Science*, 211, 453–458.

## A. Theoretical Background and Predictions

We provide a theoretical background and identification assumptions under which our between-participant comparisons in the main experiments (ONE vs TWO) are valid. Our argument builds on analyses of RIS by Azrieli et al. (2018) and Brown and Healy (2018). The identification assumption for the between-participant comparison follows Brown and Healy (2018).

### A.1. Incentive Compatibility of RIS and Monotonicity

We now introduce the setup, following Azrieli et al. (2018). Let  $R$ ,  $B$ , and  $G$  denote Bet R, Bet B, and Bet G, which appear in the main experiment, respectively. Then,  $L = \{R, B, G\}$  is the set of bets. A participant's preference for bets is represented by a complete and transitive binary relation  $\geq$  over  $L$ , with  $>$  denoting the asymmetric part. The experimenter wants to learn  $\geq$ .

In the experiment, the participant is presented with two choice situations  $D_1 = \{R, G\}$  and  $D_2 = \{B, G\}$ , from which they are asked to report a preferred bet. The list of reports in these choice situations is represented by an element of  $D = D_1 \times D_2$ .

Under RIS, the RCS is uncertain from the perspective of the participant. Let  $\Omega = \{1, 2\}$  be a state space that represents this uncertainty. The RCS is  $D_1$  in state 1 and  $D_2$  in state 2. By announcing  $(c_1, c_2) \in D$  as the most preferred bets, the participant forms a compound lottery  $\langle c_1, c_2 \rangle \in L^\Omega$  that is given by

$$\langle c_1, c_2 \rangle(\omega) = \begin{cases} c_1 & \text{if } \omega = 1, \\ c_2 & \text{if } \omega = 2. \end{cases}$$

The participant can manipulate the announcement to form a desirable compound lottery.

To describe the participant's manipulation of the announcement, we need to consider extensions  $\succeq$  of  $\geq$  over compound lotteries. An extension  $\succeq$  is a complete and transitive preference relation over  $L^\Omega$ , with  $\succ$  denoting the asymmetric part of  $\succeq$ . For each preference  $\geq$ , the experimenter has in mind a set of its extensions  $\succeq$  that are possible; such extensions are said to be admissible. The experimenter observes reports preferred by  $\succeq$  instead of  $\geq$ .

The experimenter wants the participant to report bets preferred by  $\geq$  truthfully. We say a report  $c^* \in D$  is truthful for  $\geq$  if for each  $i = 1, 2$  and  $c'_i \in D_i$ ,  $c_i^* \geq c'_i$ . RIS is incentive compatible if a report is generated by  $\succeq$  precisely when it is truthful.

**Definition 1.** RIS is incentive compatible if for any preference  $\succeq$ , any admissible extension  $\succeq$ , any  $c^* \in D$  truthful for  $\succeq$ , and any  $c' \in D$ ,  $\langle c_1^*, c_2^* \rangle \succeq \langle c'_1, c'_2 \rangle$ , with  $\langle c_1^*, c_2^* \rangle \succ \langle c'_1, c'_2 \rangle$  whenever  $c'$  is not truthful.

Azrieli et al. (2018) characterized RIS using the next condition on  $\succeq$ .

**Definition 2.**  $\succeq$  is a monotone extension of  $\succeq$  if for any  $c, c' \in D$  such that  $c_i \succeq c'_i$  for all  $i$ , we have  $\langle c_1, c_2 \rangle \succeq \langle c'_1, c'_2 \rangle$ , with  $\langle c_1, c_2 \rangle \succ \langle c'_1, c'_2 \rangle$  whenever there is some  $i$  for which  $c_i \succ c'_i$ .

We also say the participant  $(\succeq, \succeq)$  satisfies monotonicity if  $\succeq$  is a monotonic extension of  $\succeq$ . Monotonicity is a mild condition meaning that the participant prefers to make a report that yields a better bet in all states.

Azrieli et al. (2018) showed that RIS is incentive compatible under the assumption that the participant satisfies monotonicity and that RIS is essentially the only incentive-compatible mechanism when any extension that satisfies monotonicity is admissible. In other words, if the experimenter only assumes that the participant satisfies monotonicity, RIS is essentially the only incentive scheme that is expected to be incentive compatible.

## A.2. Reversal of Order and Violation of Monotonicity

Monotonicity can be violated under ambiguity, as argued by Azrieli et al. (2018). For an illustration, suppose that the participant is SAA, meaning  $G \succeq R$  and  $G \succeq B$ , and both states realize with a probability of one-half. If the participant satisfies monotonicity, he or she would have a preference  $\langle G, G \rangle \succeq \langle R, B \rangle$ , which leads to the announcement of  $(G, G)$ .

However, the participant may announce  $(R, B)$  instead of  $(G, G)$ . By announcing  $(R, B)$ , they would form a compound lottery  $\langle R, B \rangle$ , which yields  $R$  and  $B$  with the probability of one-half, as depicted in Fig. 5 (a). This compound lottery resolves in the following order: first the objective uncertainty is resolved, then the ambiguous subjective uncertainty is resolved. As Raiffa (1961) argued, the participant can identify the lottery as the one where the order of uncertainty resolution is reversed, as depicted in Fig. 5 (b). This indifference to the timing of uncertainty resolution is termed reversal of order (Anscombe and Aumann, 1963). Reversal of order implies that the participant values the compound lottery  $\langle R, B \rangle$  as highly as the one that wins with the probability of one half, irrespective of the color of the drawn ball. As the latter one stochastically dominates  $\langle G, G \rangle$ , the participant may have the preference  $\langle R, B \rangle \succ \langle G, G \rangle$ , which violates monotonicity.

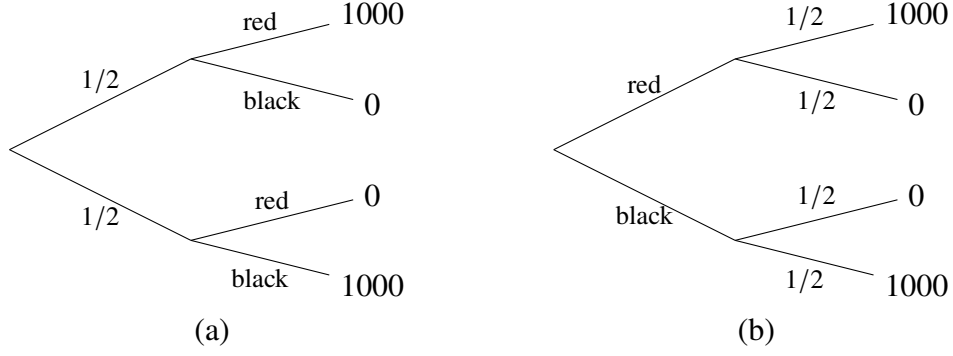


Fig. 5. Compound lotteries.

### A.3. Prediction

Our experiments were intended to evaluate whether participants satisfy monotonicity by testing predictions that follow from it. We assume that all participants satisfy the next condition, which is weaker than monotonicity.

**Assumption 1.**  $(\geq, \succeq)$  satisfies consistency if for any  $c_i, c_j \in L$ ,  $c_i \geq c_j$  if and only if  $\langle c_i, c_i \rangle \succeq \langle c_j, c_j \rangle$ .

That is, the participant satisfies consistency if his or her preferences for degenerate compound lotteries are based on those for bets. This is a weak assumption without which the experimenter cannot expect to learn  $\geq$  by observing announcements induced by  $\succeq$ . Below, the participant is assumed to satisfy consistency throughout.

As argued by Brown and Healy (2018), the participant's preference  $\succeq$  can vary according to the framing of the experiment. Thus, the performance of RIS should be evaluated under the control of the framing, which is why we designed treatment ONE as close to treatment TWO as possible.

If the difference in incentive schemes also has a framing effect on  $\succeq$ , our comparisons do not work. We assume the next condition, which follows Brown and Healy (2018).

**Assumption 2** (Mechanism Invariance).  $\succeq$  satisfies mechanism invariance if  $\succeq$  does not differ between two experiments that are identical except for their incentive schemes.

We proceed to make a prediction tested by our experiment.

**Prediction 1.** Assume that all participants satisfy consistency and mechanism invariance. If all participants further satisfy monotonicity, then the proportion of participants who choose  $G$  in

Question R (resp. Question B) is the same for those in ONE with incentivized Question R (resp. Question B) and those in TWO.

Although the derivation of this prediction follows Brown and Healy (2018), we reproduce it here for completeness. First, we argue that, under monotonicity, a participant’s answer to the incentivized question in ONE is the same as that to the question in TWO. Imagine that a participant is in ONE, and that Question R is incentivized. Suppose further that he or she answers that question by choosing Bet G over Bet R.<sup>13</sup> Because Question R is the only incentivized question, this answer reveals that  $\langle G, G \rangle \succeq \langle R, R \rangle$ . Then, using the assumption of consistency, we infer that  $G \succeq R$  in ONE. As the only difference between ONE and TWO is in the incentive schemes, from this preference and the assumption of mechanism invariance, we infer that the participant would also have preference  $G \succeq R$  in TWO. Then, using the assumption of monotonicity, we infer that he or she would answer Question R by choosing Bet G over Bet R if they were in TWO.

As we randomly assign participants to ONE and TWO, the distributions of participants’ preferences among these treatments are considered to be the same. Thus, under monotonicity, we can predict that the proportion of participants who choose Bet G in Question R is the same between participants in ONE with incentivized Question R and those in TWO.

## B. Instructions for the experiment

For the replication experiment of Baillon et al. (2022a), we translated their original instructions into Japanese and adjusted the reward as described in the main text. Because the original instructions in Baillon et al. (2022a) do not mention a participation fee, we added the following statement orally at the end of the instructions to be consistent with our main experiment: “If you finish this experiment, we will pay you 500 yen as a participation fee.” The Japanese version of the instructions used in our replication of Baillon et al. (2022a) is available from the authors upon request.

For our main experiment, the following are the instructions translated from Japanese.

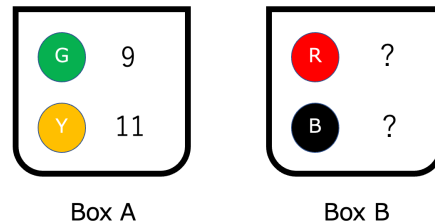
---

<sup>13</sup>A similar argument applies to the case where Question B is incentivized or the participant reports preference for an ambiguous bet.

## B.1. Instructions for ONE

Thank you for participating in this experiment. If you finish this experiment, we will pay you 500 yen as a participation fee. Depending on the results of the decisions you make during the experiment, you may be paid 1000 yen in addition to the participation fee.

There are two boxes. Box A contains 9 green balls and 11 yellow balls, for a total of 20. Box B contains a total of 20 red and black balls. The number of red and black balls in Box B was decided by a participant in another experiment that was conducted recently. We will not tell you the number of red and black balls.



Later, you will be asked to indicate your preferences for the following three lotteries.

- Lottery Green: If a ball taken out of **Box A** is **green**, you are paid 1000 yen, but if it is yellow, you are not paid anything.
- Lottery Red: If a ball taken out of **Box B** is **red**, you are paid 1000 yen, but if it is black, you are not paid anything.
- Lottery Black: If a ball taken out of **Box B** is **black**, you are paid 1000 yen, but if it is red, you are not paid anything.

There are two questions.

- Question (Red): Which do you prefer, Lottery Green or Lottery Red?
- Question (Black): Which do you prefer, Lottery Green or Lottery Black?

You will later receive one answer sheet for each question. Please write your answers to the questions on those sheets.

## Procedure for Reward Payment

Additional rewards will be paid based on the answers to Question (Red). After the experimenter collects the answer sheet, the assistant will draw one ball from Box A and one from Box B *without looking inside the boxes*. The assistant will announce the colors of both balls and record them on the whiteboard.

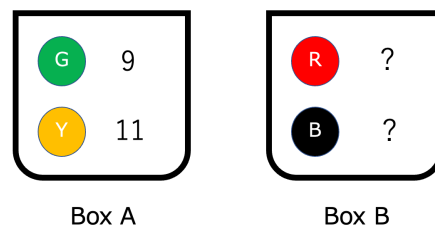
The rewards will be paid as follows.

- If you answer Question (Red) by choosing Lottery Green, you win if the assistant draws a green ball from Box A (45%).
- If you answer Question (Red) by choosing Lottery Red, you win if the assistant draws a red ball from Box B.

## B.2. Instructions for TWO

Thank you for participating in this experiment. If you finish this experiment, we will pay you 500 yen as a participation fee. Depending on the results of the decisions you make during the experiment, you may be paid 1000 yen in addition to the participation fee.

There are two boxes. Box A contains 9 green balls and 11 yellow balls, for a total of 20. Box B contains a total of 20 red and black balls. The number of red and black balls in Box B was decided by a participant in another experiment that was conducted recently. We will not tell you the number of red and black balls.



Later, you will be asked to indicate your preferences for the following three lotteries.

- Lottery Green: If a ball taken out of **Box A** is **green**, you are paid 1000 yen, but if it is yellow, you are not paid anything.
- Lottery Red: If a ball taken out of **Box B** is **red**, you are paid 1000 yen, but if it is black, you are not paid anything.

- Lottery Black: If a ball taken out of **Box B** is **black**, you are paid 1000 yen, but if it is red, you are not paid anything.

There are two questions.

- Question (Red): Which do you prefer, Lottery Green or Lottery Red?
- Question (Black): Which do you prefer, Lottery Green or Lottery Black?

You will later receive one answer sheet for each question. Please write your answers to the questions on those sheets.

### **Procedure for Reward Payment**

Additional rewards will be paid based on the answers to the two questions. To decide which question is the target of the rewards, the assistant will roll a six-sided die once, with three sides painted red and three sides painted black. The question corresponding to the color of the roll will be the target of the rewards for all participants. You will be given an envelope containing a piece of paper with the reward question on and will write your participant ID on it. *Do not open the envelope until you are told to do so.* Note that the question that determines the final payment amount is written on the paper inside the envelope and is determined *before* you answer the question. After you answer the questions, the experimenter will tell you to open the envelope. The assistant will then draw one ball from Box A and one ball from Box B *without looking inside the boxes*. The assistant will announce the color of both balls and record it on the whiteboard.

Based on the answers to the questions written on the paper inside the envelope, you will receive rewards as follows.

If Question (Red) is selected:

- In you choose Lottery Green, you win if the assistant draws the green ball from Box A (45%).
- In you choose Lottery Red, you win if the assistant draws a red ball from Box B.

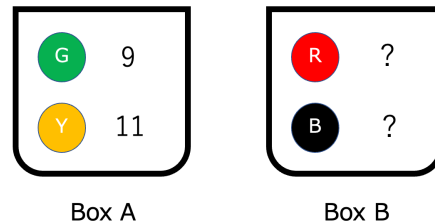
If Question (Black) is selected:

- In you choose Lottery Green, you win if the assistant draws a green ball from Box A (45%).
- In you choose Lottery Black, you win if the assistant draws a black ball from Box B.

### B.3. Instructions for SINGLE

Thank you for participating in this experiment. If you finish this experiment, we will pay you 500 yen as a participation fee. Depending on the results of the decisions you make during the experiment, you may be paid 1000 yen in addition to the participation fee.

There are two boxes. Box A contains 9 green balls and 11 yellow balls, for a total of 20. Box B contains a total of 20 red and black balls. The number of red and black balls in Box B was decided by a participant in another experiment that was conducted recently. We will not tell you the number of red and black balls.



Later, you will be asked to indicate your preferences for the following three lotteries.

- Lottery Green: If a ball taken out of **Box A** is **green**, you are paid 1000 yen, but if it is yellow, you are not paid anything.
- Lottery Red: If a ball taken out of **Box B** is **red**, you are paid 1000 yen, but if it is black, you are not paid anything.
- Lottery Black: If a ball taken out of **Box B** is **black**, you are paid 1000 yen, but if it is red, you are not paid anything.

There are two questions.

- Question (Red): Which do you prefer, Lottery Green or Lottery Red?
- Question (Black): Which do you prefer, Lottery Green or Lottery Black?

You will later receive one answer sheet for each question. You will answer only one question. First, please choose which question you would like to answer. On each answer sheet, you will be asked if you wish to answer that question. On the answer sheet for the question you decide to answer, write your answer to that question.

## Procedure for Reward Payment

Additional rewards will be paid based on the answers to a question. After the experimenter collects the answer sheet, the assistant will draw one ball from Box A and one from Box B *without looking inside the boxes*. The assistant will announce the colors of both balls and record them on the whiteboard.

Based on the answers to the questions you choose, you will receive rewards as follows.

If you choose Question (Red):

- If you answered Question (Red) by choosing Lottery Green, you win if the assistant draws a green ball from Box A (45%).
- If you answered Question (Red) by choosing Lottery Red, you win if the assistant draws a red ball from Box B.

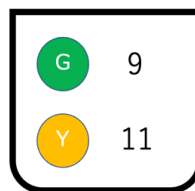
If you choose Question (Black):

- If you answered Question (Black) by choosing Lottery Green, you win if the assistant draws a green ball from Box A (45%).
- If you answered Question (Black) by choosing Lottery Black, you win if the assistant draws a black ball from Box B.

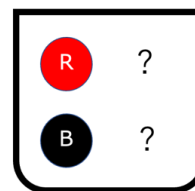
## C. Answer sheet for ONE and TWO (English translation)

Answer sheet for Question (Red)

Please write your participant ID in the box below



Box A



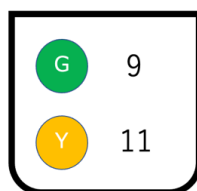
Box B

Please circle a or b

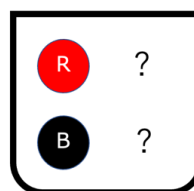
- a) Lottery Green: If a ball taken out of **Box A** is **green**, you are paid 1000 yen, but if it is yellow, you are not paid anything.
- b) Lottery Red: If a ball taken out of **Box B** is **red**, you are paid 1000 yen, but if it is black, you are not paid anything.

Answer sheet for Question (Black)

Please write your participant ID in the box below



Box A



Box B

Please circle a or b

- a) Lottery Green: If a ball taken out of **Box A** is **green**, you are paid 1000 yen, but if it is yellow, you are not paid anything.
- b) Lottery Black: If a ball taken out of **Box B** is **black**, you are paid 1000 yen, but if it is black, you are not paid anything.

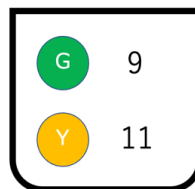
## D. Answer sheet for SINGLE (English translation)

Answer sheet for Question (Red)

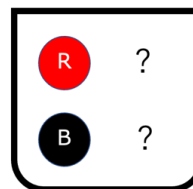
Please write your participant ID in the box below

Please circle a or b

- a) Answer this question.
- b) Do not answer this question.



Box A



Box B

Please circle a or b only if you are answering this question

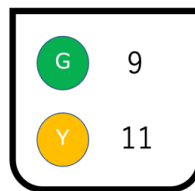
- a) Lottery Green: If a ball taken out of **Box A** is **green**, you are paid 1000 yen, but if it is yellow, you are not paid anything.
- b) Lottery Red: If a ball taken out of **Box B** is **red**, you are paid 1000 yen, but if it is black, you are not paid anything.

Answer sheet for Question (Black)

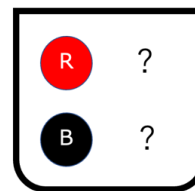
Please write your participant ID in the box below

Please circle a or b

- a) Answer this question.
- b) Do not answer this question.



Box A



Box B

Please circle a or b only if you are answering this question

- a) Lottery Green: If a ball taken out of **Box A** is **green**, you are paid 1000 yen, but if it is yellow, you are not paid anything.
- b) Lottery Black: If a ball taken out of **Box B** is **black**, you are paid 1000 yen, but if it is black, you are not paid anything.

# Online Supplementary Material for Experimental Evaluation of Random Incentive System under Ambiguity

Tomohito Aoyama Nobuyuki Hanaki

For completeness, we report the designs, implementations, and results of a series of experiments we have reported in the earlier version of the paper in this online supplementary material. We also report English translations of the instructions for these experiments (except for the preliminary laboratory experiment as they are the same, except for the amount of the participation fee (1000 JPY instead of 500 JPY) and the additional reward (500 JPY instead of 1000 JPY) reported in the internal appendix of the paper.)

## I. Online Experiment 1

As in our main experiment reported in Section 2.1, participants were presented with two boxes, Box A and Box B, as depicted in Figure 1, and were presented with three bets –Bet G, Bet R, and Bet B– with a smaller reward (500 JPY instead of 1000 JPY). In addition, they were asked to answer the same two questions R and B as in the main experiment.

The two questions were displayed on separate individual pages, but as participants received instructions about the two questions before making their choices, they could view the entire experiment as a single decision problem. At the beginning of the experiment, either Question R or Question B was selected as the target of rewards, or the RCS.<sup>1</sup> The method of selecting the RCS varied across treatments. After participants made choices, one ball was drawn from each of Box A and Box B, which were simulated on the computer. Then, participants received payments based on the outcome of the bet they chose in the RCS.

### I.1. Treatments

There were two treatments, ONE and TWO, where the selection procedures of the RCS differed.

---

<sup>1</sup>In all the experiments, we selected the RCS under RIS at the beginning, which might enhance the incentive compatibility of RIS, as argued by Baillon et al. (2022b).

Table I.1  
Groups of participants in the main experiment.

|         | Session 1 | Session 2 |
|---------|-----------|-----------|
| Group A | ONE       | TWO       |
| Group B | TWO       | ONE       |
| Group C | ONE       | ONE       |

In ONE, the RCS was the question displayed later. For instance, if Question R appeared first and then Question B appeared, the RCS would be Question B. Participants were informed of this selection method. The order of the questions was randomized across participants, resulting in Question R being the RCS for half of the participants and Question B for the other half. We informed participants about this randomization to prevent them from suspecting that we were presenting a bet with a lower probability of winning in the second question.

In TWO, the RCS was selected following RIS. Participants were informed that the computer randomly chose each question as the RCS with a probability of one half before answering the questions. The RCS was revealed after participants answered both questions.

## I.2. Sessions

Participants were not informed beforehand that there would be two sessions. We recruited those who participated in Session 1 for Session 2 and connected the data using their ID number. Participants were randomly assigned to one of three groups that differed in the treatments in the two sessions, as summarized in Table I.1. Group A was assigned to ONE in Session 1 and to TWO in Session 2, while Group B was assigned to TWO in Session 1 and ONE in Session 2. Group C was assigned to ONE in both sessions.

The numbers of red and black balls in the two sessions were determined by different students. In Session 2, we informed participants of this fact to prevent them from making their decision based on the outcome of Session 1. While the order of Question R and Question B was randomized among participants, it remained fixed across the two sessions.

## I.3. Main Variable

The variable of interest is the answer to the question displayed later. In ONE, participants were expected to report their preferences truthfully in response to that question because it was known to be the RCS. In TWO, participants might manipulate answers to hedge ambiguity through the

randomness of RIS. If participants in these treatments answered the second question differently, we conclude that the use of RIS may have been distorting their choices.

Both between- and within-participant comparisons are conducted. For the between-participant comparison, we compare the proportions of participants who chose Bet G in the second question between ONE and TWO. For the within-participant comparison, Group C serves as a benchmark. Participants in Group C were assigned to ONE in both sessions. Note that it is possible that they answer the second question differently between sessions for some reason. However, participants in Groups A and B were assigned to different treatments in the two sessions, so their choices may differ because of the difference in the incentive schemes, in addition to stochasticity in choices as in Group C. If the proportion of participants in Groups A and B who answer differently in the two sessions is greater than that of Group C, we can conclude that the difference in incentive schemes caused participants to change their answers.

## **I.4. Results**

The experiment was conducted online in two sessions, two weeks apart, in December 2021 using Qualtrics ([www.qualtrics.com](http://www.qualtrics.com)). Invitation emails were sent to students in the subject pool of Institute of Social and Economic Research (ISER), Osaka University, managed by ORSEE (Greiner, 2015). Among them, 413 participants completed Session 1, of which 312 also completed Session 2. Participants were instructed to complete the task individually by clicking on the link they received via email on the same day. On average, they took 156 seconds and 155 seconds to complete Sessions 1 and 2, respectively. On average, they earned 329 JPY and 326 JPY in Sessions 1 and 2, respectively, including a participation fee of 100 JPY each. The reward was an Amazon gift card via email.<sup>2</sup>

### **I.4.1. Sample Attrition**

Because we invited participants to Session 2 after they completed Session 1, there was some attrition of the sample. To examine the effect of the treatment and the outcome participants experienced in Session 1 on the rate of attrition, we conducted regression analyses to investigate the determinants of participation in Session 2. Table I.2 shows the regression results.

The dependent variable is a dummy that takes the value of one if a participant proceeded to

---

<sup>2</sup>This experiment was pre-registered on October 30, 2021: [https://aspredicted.org/DFY\\_K5Y](https://aspredicted.org/DFY_K5Y).

Session 2 and zero otherwise. The independent variables, all of which are dummies that take values of one or zero, are as follows.

- Green takes the value of one if participants answered Bet G is preferred to the second question of Session 1.
- Win takes the value of one if they won in Session 1.
- TWO takes the value of one if they were involved in TWO in Session 1.
- Order takes the value of one if the first question was Question B.

As can be seen from Table I.2, none of these variables are significant at even the 10% significance level. Thus, we conclude there is no significant effect of the treatment or the outcome of Session 1 on participation in Session 2.

Furthermore, we tested several types of order effects. First, in each session, the order of questions did not affect answers to the second question. The  $p$ -values are  $p = 0.46$  and  $p = 1$  (resp.  $p = 0.15$  and  $p = 0.96$ ) for ONE and TWO in Session 1 (resp. in Session 2), respectively. Second, the order of treatments across the two sessions did not affect the answers to the second question in each treatment. The  $p$ -values are  $p = 0.67$  and  $p = 0.25$  for ONE and TWO, respectively. Overall, we found no significant order effect.

#### **I.4.2. Effect of RIS**

Because there was no observed effect of the treatment or the outcome in Session 1 on the rate of attrition, we pool the data from the three groups when possible. In Session 1, 257 participants were involved in ONE and 156 in TWO. In Session 2, 191 participants were involved in ONE and 121 in TWO.

The  $\chi^2$  test (with Yate's continuity correction, which is default in R) was used unless otherwise noted. Fisher's exact test was also used, but the results are not reported as they are similar to those obtained from the  $\chi^2$  test.<sup>3</sup> A significance level of 5% was adopted for all tests.

We first report the result of the between-participant comparison. In Session 1, 64.6% and 68.6% of participants chose Bet G in the second question in ONE and TWO, respectively. This difference

---

<sup>3</sup>We report the  $p$ -values based on the Fisher's exact test in footnotes for completeness.

Table I.2  
Determinants of participation in Session 2.

|                         | Dependent variable:                          |                     |
|-------------------------|----------------------------------------------|---------------------|
|                         | Participation                                |                     |
|                         | OLS<br>(1)                                   | logistic<br>(2)     |
| Green                   | 0.012<br>(0.045)                             | 0.067<br>(0.243)    |
| Win                     | 0.070<br>(0.044)                             | 0.387<br>(0.240)    |
| TWO                     | 0.018<br>(0.044)                             | 0.098<br>(0.239)    |
| Order                   | 0.012<br>(0.043)                             | 0.062<br>(0.235)    |
| Constant                | 0.702***<br>(0.049)                          | 0.849***<br>(0.257) |
| Observations            | 413                                          | 413                 |
| R <sup>2</sup>          | 0.007                                        |                     |
| Adjusted R <sup>2</sup> | −0.002                                       |                     |
| Log Likelihood          |                                              | −228.187            |
| Akaike Inf. Crit.       |                                              | 466.374             |
| Residual Std. Error     | 0.431 (df = 408)                             |                     |
| F Statistic             | 0.769 (df = 4; 408)                          |                     |
| Note:                   | * $p < 0.1$ ; ** $p < 0.05$ ; *** $p < 0.01$ |                     |

is not statistically significant ( $p = 0.47$ ).<sup>4</sup> A similar analysis using the data from Session 2 also shows no significant difference ( $p = 0.91$ ).<sup>5</sup>

**Result I.1** (Between-Participant Comparison). The use of RIS did not have a significant effect on choices.

We next proceed to the within-participant comparison. Of participants in Groups A and B,

<sup>4</sup>This analysis uses the data of all participants, including those who did not proceed to Session 2. Therefore, the sample sizes for ONE and TWO are 257 and 156, respectively. The  $p$ -value based on the Fisher's exact test is 0.453.

<sup>5</sup>The proportions of participants who chose Bet G were 62.3% (119 out of 191) and 63.6% (77 out of 121) in ONE and TWO, respectively. The  $p$ -value based on the Fisher's exact test is 0.904.

31.3% changed their answers, while 30.6% of those in Group C did. These proportions were not significantly different ( $p = 0.973$ ).<sup>6</sup>

**Result I.2** (Within-Participant Comparison). The assignment to different treatments between sessions did not have a significant effect on their propensity to change their answers.

These results suggest that RIS is incentive compatible.

## II. Online Experiment 2

The second online experiment aimed to investigate the impact of using a benchmark treatment similar to Baillon et al. (2022a) and the effect of the way control treatment is framed in our online experiment 1. The tasks were similar to those in the main experiment, but the incentive schemes and framings differed.

### II.1. Treatments

There were two treatments, SINGLE and ONE\*, with differing incentive schemes.

SINGLE was designed to resemble the control treatment Single of Baillon et al. (2022a). Participants in SINGLE first selected either Question R or Question B as the RCS. They were then shown their selected question and chose a bet.

ONE\* was a variation of ONE in the main experiment, differing only in the framing. In ONE\*, participants were provided with only one question, which was the RCS. As in ONE, we informed participants in ONE\* that the RCS was Question R for half of them and Question B for the other half.

### II.2. Main Variable

The variable of interest is the answer to the displayed question. We also compare the observed ambiguity attitudes using data from SINGLE in this experiment and TWO in online experiment 1. Participants in SINGLE are classified as SAA if they chose Bet G in the RCS. Participants in TWO are classified as SAA if they chose Bet G in response to both questions. The proportions of SAA participants in these treatments are compared.

---

<sup>6</sup>This analysis uses the data of participants who completed Session 2. The sample size for Groups A and B is 240, and that for Group C is 72. The  $p$ -value from the Fisher's exact test is 1.000.

Additionally, we examine the effect of the way control treatment is framed in our main experiment. For this purpose, we compare the proportion of participants who chose Bet G in ONE\* in this experiment with the proportion of participants who chose Bet G in the second question in TWO of online experiment 1.

### II.3. Results

The experiment was conducted online in January 2022 using Qualtrics. Invitation emails were sent to students in the same subject pool as in online experiment 1.<sup>7</sup> Of the invited students, 206 completed the experiment. Participants were asked to complete the task individually by clicking on the individualized link they received via email on the same day. On average, participants took 124 seconds to complete the experiment. On average, they earned 360 JPY, including 100 JPY as a participation fee. Participants received their reward in the form of an Amazon gift card by email.<sup>8</sup>

The number of participants in ONE\* and SINGLE were 103 each. We first compare SINGLE and TWO. In SINGLE, the proportion of SAA participants was 58.3%. In TWO, it was 58.3% in Session 1 and 53.7% in Session 2. There was no significant difference in the proportions of SAA participants between SINGLE and TWO ( $p = 0.91$  for Session 1 and  $p = 0.59$  for Session 2).<sup>9</sup>

**Result II.1** (Proportion of SAA participants). The use of RIS did not have a significant effect on the observed ambiguity aversion.

We next compare ONE\* v.s. TWO and ONE. As reported in Section I, 68.6% and 64.6% of participants in TWO and ONE of Session 1 answered Bet G is preferred in response to the second question. However, 52.4% of participants in ONE\* chose Bet G. These differences are significant ( $p = 0.013$  (vs TWO) and  $p = 0.043$  (vs ONE)).<sup>10</sup>

**Result II.2** (Comparison of ONE\* v.s. TWO and ONE). Participants were less likely to choose Bet G in ONE\* compared with TWO and ONE

That is, using ONE\* instead of ONE allows us to conclude that the use of RIS enhances SAA behavior, which is the opposite effect of that reported in Baillon et al. (2022a). This result suggests that the framing of control treatment in the main experiment could influence the results.

---

<sup>7</sup>Participants in online experiment 1 were excluded from the invitation.

<sup>8</sup>This experiment was pre-registered on January 12, 2022: [https://aspredicted.org/3J1\\_Y7T](https://aspredicted.org/3J1_Y7T).

<sup>9</sup>The  $p$ -values based on the Fisher's exact test are 1.000 for Session 1 and 0.503 for Session 2.

<sup>10</sup>The  $p$ -values based on the Fisher's exact test are 0.013 (vs TWO) and 0.042 (vs ONE).

### III. Preliminary Laboratory Experiment

The third experiment was intended to examine the overall effect of experimenting in a laboratory and following PRINCE, the guideline followed in the experiment of Baillon et al. (2022a). The task, treatments, and main variables were the same as those reported in Section 2.1 of the main text, but with the additional earning from winning the bet set at 500 JPY (instead of 1000 JPY), as in the two online experiments, and the participation fee is set to 1000 JPY (instead of 500 JPY).

This experiment was conducted at the experimental laboratory of Osaka University in July 2022. A total of 126 participants (students of Osaka University, which is the same subject pool as online experiments 1 and 2) participated in the experiment in seven sessions, excluding randomly selected assistants.<sup>11</sup> The average session duration was 20 minutes. On average, participants earned 1234 JPY, including 1000 JPY as a participation fee.<sup>12</sup>

#### III.1. Results

The number of participants in ONE-R and ONE-B were 20 and 21, respectively. Those in TWO and SINGLE were 41 and 44, respectively. We first compare the data from ONE and TWO. The proportion of participants who preferred Bet G to Bet R was 75% in ONE-R and 44% in TWO. The proportion of participants who preferred Bet G to Bet B was 33% in ONE-B and 51% in TWO. The  $p$ -values for these differences are  $p = 0.044$  and  $p = 0.285$ , respectively.<sup>13</sup> With the Bonferroni corrected significance level of 2.5% (as we conduct multiple comparisons), the differences are not significant.

**Result III.1** (Between-Participant Comparison in Lab). The use of RIS did not have a significant effect on choices.

The propensity to choose the risky bet (Bet G) over an ambiguous bet (Bet R or Bet B) in the RCS is significantly greater in ONE-R than in ONE-B ( $p = 0.018$ ). This observation cannot be

---

<sup>11</sup>Participants in online experiments 1 and 2 were excluded from the invitation.

<sup>12</sup>This experiment was pre-registered on July 9, 2022: [https://aspredicted.org/3N4\\_MB8](https://aspredicted.org/3N4_MB8)

<sup>13</sup>Note that we do not know in which order participants in the preliminary laboratory experiment answered the questions because participants received an answer sheet that has two sides, while in the online experiment 1, we know (because we know in which order these two questions appeared on the screen). Thus, in the online experiment 1, we compared the choice in the second question participants answered (which was the RCS in ONE) in treatments ONE and TWO. Because we cannot do the same for the preliminary laboratory experiment, we do not pool the data, and compare separately ONE-R and ONE-B with answers in questions R and B, respectively, in TWO. The  $p$ -values based on the Fisher's exact test are 0.03 and 0.281, respectively.

explained by a bias in participants' belief that Box U contained many black balls because this is not consistent with the data from TWO and SINGLE. In fact, participants in TWO did not have a different propensity to choose Bet G in the two questions ( $p = 0.37$ , McNemar test). Furthermore, the number of participants in SINGLE who chose Bet R and Bet B is the same. We could not find an explanation for the difference between ONE-R and ONE-B (other than possibly an error).

We next compare SINGLE and TWO. The proportion of SAA participants is 50% in SINGLE and 34.1% in TWO. This difference is not significant ( $p = 0.21$ ).<sup>14</sup>

**Result III.2** (Proportion of SAA participants in Lab). The use of RIS did not have a significant effect on the observed ambiguity aversion.

## IV. Instructions for Online Experiment 1

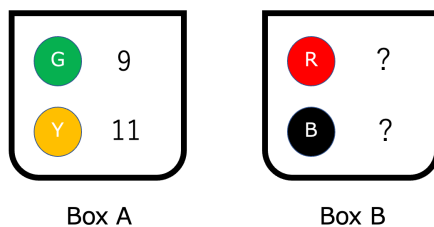
The following are the instructions translated from Japanese. Because the introduction is common to all treatments, we only show the introduction for ONE.

### IV.1. Instruction for ONE

#### Introduction

Thank you for participating in this experiment. If you finish this experiment, we will pay you 100 yen as a participation fee. Depending on the results of the decisions you make during the experiment, you may be paid 500 yen in addition to the participation fee. Please note that we will not be able to pay you if your response time for the entire experiment exceeds 30 minutes.

#### Experiment description




---

<sup>14</sup>The  $p$ -value based on the Fisher's exact test is 0.188.

Two boxes are simulated on a computer. Both boxes contain 20 balls. Box A contains 9 green balls and 11 yellow balls, for a total of 20. Box B contains a total of 20 red and black balls. The number of red and black balls in Box B was decided by a participant in another experiment that was conducted recently. We will not tell you the numbers of red and black balls.

Later, you will be asked to indicate your preferences for the following three lotteries.

- Lottery Green: If a ball taken out of **Box A** is **green**, you are paid 500 yen, but if it is yellow, you are not paid anything.
- Lottery Red: If a ball taken out of **Box B** is **red**, you are paid 500 yen, but if it is black, you are not paid anything.
- Lottery Black: If a ball taken out of **Box B** is **black**, you are paid 500 yen, but if it is red, you are not paid anything.

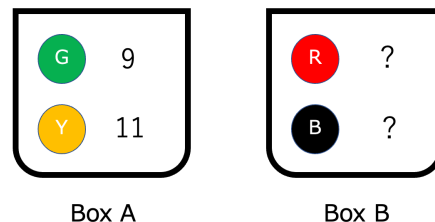
There are two questions. In both questions, we will present two of the three lotteries and ask you to choose one of them.

- Question 1: Which do you prefer, Lottery Green or Lottery Red?
- Question 2: Which do you prefer, Lottery Green or Lottery Black?

*After these two questions, you will receive the lottery ticket you choose in Question 2. Then, a ball will be drawn from a box simulated on the computer, and if the lottery ticket is a winner, you will receive 500 yen in addition to the participation fee.*

Half of the participants in this experiment will receive what they choose from Lottery Green and Lottery Red, and the other half will receive what they choose from Lottery Green and Lottery Black.

## IV.2. Instruction for TWO



Two boxes are simulated on a computer. Both boxes contain 20 balls. Box A contains 9 green balls and 11 yellow balls, for a total of 20. Box B contains a total of 20 red and black balls. The number of red and black balls in Box B was decided by a participant in another experiment that was conducted recently. We will not tell you the number of red and black balls.

Later, you will be asked to indicate your preferences for the following three lotteries.

- Lottery Green: If a ball taken out of **Box A** is **green**, you are paid 500 yen, but if it is yellow, you are not paid anything.
- Lottery Red: If a ball taken out of **Box B** is **red**, you are paid 500 yen, but if it is black, you are not paid anything.
- Lottery Black: If a ball taken out of **Box B** is **black**, you are paid 500 yen, but if it is red, you are not paid anything.

There are two questions. In both questions, we will present two of the three lotteries and ask the participants to choose one of them.

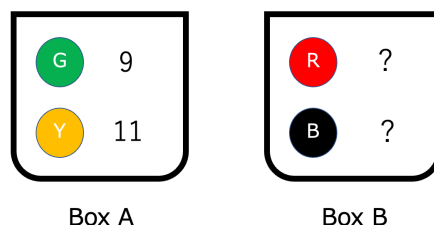
- Question 1: Which do you prefer, Lottery Green or Lottery Red?
- Question 2: Which do you prefer, Lottery Green or Lottery Black?

At the beginning of the experiment, the computer will randomly choose either Question 1 or Question 2 as the target of the reward. Both questions will be chosen with a probability of one-half. We will not tell you which question is the target of the reward until the reward payment stage. You will receive a lottery ticket for your choice on the question that is the target of the reward. Then, a ball will be drawn from the computer-simulated box, and if the lottery you receive is a winner, you will receive 500 yen in addition to your participation fee.

## V. Instructions for Online Experiment 2

The following are the instructions translated from Japanese. The instructions included an introduction that is similar to the introduction to Online Experiment 1, which is omitted here.

## V.1. Instruction for SINGLE



Two boxes are simulated on a computer. Both boxes contain 20 balls. Box A contains 9 green balls and 11 yellow balls, for a total of 20. Box B contains a total of 20 red and black balls. The number of red and black balls in Box B was decided by a participant in another experiment that was conducted recently. We will not tell you the number of red and black balls.

Later, you will be asked to indicate your preferences for the following three lotteries.

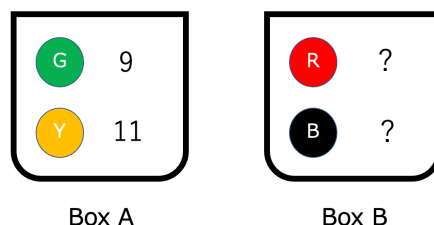
- Lottery Green: If a ball taken out of **Box A** is **green**, you are paid 500 yen, but if it is yellow, you are not paid anything.
- Lottery Red: If a ball taken out of **Box B** is **red**, you are paid 500 yen, but if it is black, you are not paid anything.
- Lottery Black: If a ball taken out of **Box B** is **black**, you are paid 500 yen, but if it is red, you are not paid anything.

First, you will be asked which of the following questions you would like to answer.

- Question (Red): Which do you prefer, Lottery Green or Lottery Red?
- Question (Black): Which do you prefer, Lottery Green or Lottery Black?

If you respond that you want to answer Question (Red), Question (Red) will be displayed. The same applies to Question (Black). *After you answer the displayed question, you will receive the lottery ticket that you choose.* Then, a ball will be drawn from the computer-simulated box, and if the lottery you receive is a winner, you will receive 500 yen in addition to your participation fee.

## V.2. Instruction for ONE\*



Two boxes are simulated on a computer. Both boxes contain 20 balls. Box A contains 9 green balls and 11 yellow balls, for a total of 20. Box B contains a total of 20 red and black balls. The number of red and black balls in Box B was decided by a participant in another experiment that was conducted recently. We will not tell you the number of red and black balls.

Later, you will be asked to indicate your preferences for two of the following three lotteries.

- Lottery Green: If a ball taken out of **Box A** is **green**, you are paid 500 yen, but if it is yellow, you are not paid anything.
- Lottery Red: If a ball taken out of **Box B** is **red**, you are paid 500 yen, but if it is black, you are not paid anything.
- Lottery Black: If a ball taken out of **Box B** is **black**, you are paid 500 yen, but if it is red, you are not paid anything.

There is only one question posed, as follows.

Question: Which lottery will you choose, Lottery Green or Lottery Black?

*After you answer the question, you will receive the lottery ticket you choose.* Then, a ball will be drawn from the computer-simulated box, and if the lottery you receive is a winner, you will receive 500 yen in addition to your participation fee.

Half of the participants in this experiment will be presented with Lottery Green and Lottery Red, and the other half will be presented with Lottery Green and Lottery Black.