DOES EYE-TRACKING HAVE AN EFFECT ON ECONOMIC BEHAVIOR?

Jennifer Kee
Melinda Knuth
Joanna Lahey
Marco A. Palma

Does Eye-Tracking Have an Effect on Economic Behavior?
Jennifer Kee, Melinda Knuth, Joanna Lahey, and Marco A. Palma
NBER Working Paper No. 28223
December 2020
JEL No. C9,D03,D8

ABSTRACT

Eye-tracking is becoming an increasingly popular tool for understanding the underlying behavior driving economic decisions. However, an important unanswered methodological question is whether the use of an eye-tracking device itself induces changes in the behavior of experiment participants. We study this question using eight popular games in experimental economics. We implement a simple design where participants are randomly assigned to either a control or an eye-tracking treatment condition. In seven of the eight games, eye-tracking did not produce different outcomes. In the Holt and Laury risk assessment (HL), subjects with multiple calibration attempts behave like outliers under eye-tracking conditions, skewing the overall results. Further exploration shows that poor calibrators also show marginally higher levels of negative emotion, which is correlated with higher risk aversion in both HL and in the Eckel and Grossman gambling tasks. Because difficulty calibrating is correlated with eye-tracking data quality, the standard practice of removing participants who did not have good eye-tracking data quality resulted in no difference between the treatment and control groups in HL. Our results suggest that experiments may incorporate eye-tracking equipment without inducing changes in the economic behavior of participants, particularly after observations with low eye-tracking quality are removed.

Jennifer Kee
Texas A&M University
600 John Kimbrough Blvd, 2124 TAMU
College Station, TX 77843-2124
USA
young09@tamu.edu

Melinda Knuth
University of Florida
2725 S. Binion Road
Apopka, FL 32703
melindaknuth@ufl.edu

Joanna Lahey
The Bush School
Texas A&M University
Mailstop 4220
College Station, TX 77843
and NBER
jlahey@tamu.edu

Marco A. Palma
Texas A&M University
1500 Research Pkwy
College Station, TX 77843
mapalma@tamu.edu

1 Introduction

The application of eye-tracking technology to investigate human behavior is gaining popularity in economics (Lahey and Oxley, 2016). Eye-tracking has proved to be a versatile methodological tool to evaluate intrinsic human responses (Fengs, 2011; Knoepfle et al., 2009; Reutskaja et al., 2011; Rosch and Vogel-Walcutt, 2013; Sharafi et al., 2015; Wang et al., 2010; Wedel, 2014). Specifically, within economic experiments, eye-tracking has been previously used to assess visual attention, pupil dilation, mental effort, and, more recently, strategic interaction during economic games (Wang et al., 2010). Eye-tracking has also been used extensively in psychology and related fields to study cognitive load, reading efficiency, complexity of information, attention, decision making under time pressure, and more (Fiedler et al., 2013; Pieters and Wedel, 2007; Reutskaja et al., 2011). Eye-tracking technology is frequently used to evaluate the nuances of behavior in game theory. With the influx of experiments conducted with eye-tracking in the past decade, the use of eye-tracking will continue to gain popularity in economics (see Lahey and Oxley (2016) for a literature review, also Sickmann and Le (2016), and Wang et al. (2010)). The beauty contest (Chen et al., 2018; Chen and Krajbich, 2017), prisoner's dilemma (Devetag et al., 2016; Fiedler et al., 2013; Hristova and Grinberg, 2005; Peshkovskaya et al., 2017; Stewart et al., 2016), leader games (Stewart et al., 2016), public goods (Tanida and Yamagishi, 2010), two-armed bandit learning (Hu et al., 2013), variations of the normal form game of the prisoner's dilemma and stag hunt (Knoepfle et al., 2009; Polonio et al., 2015; Popovic, 2014), sender-receiver (Wang et al., 2010), and coordination games (Król and Król, 2017) have been studied using eye-tracking and other biometric equipment.

Even though eye-tracking provides rich data that enables researchers to better understand individual decisions, it is possible that participants may modify their behavior when they know their eye movements are being recorded (Wang et al., 2010). People may behave in a more socially desirable way if, by monitoring their eyes, they feel pressured to conform to social expectations (Kaminska and Foulsham, 2013). Yet, the question of whether the

use of an eye-tracking device itself induces a change in experiment participant behavior has not been comprehensively studied (Lejarraga et al., 2017; Wang et al., 2010). In this paper, we investigate whether the use of an eye-tracking device induces changes in the economic behavior in incentivized laboratory experiments. If it does, external validity would be limited to situations in which participants know they are being watched. We also study if any such effects go away over time. If they do, then later eye-tracking results can be used. We address these issues using eight standard economic games in which participants are randomly assigned to either an eye-tracking treatment or no eye-tracking control. Social desirability is the tendency of some respondents to report an answer in a way they deem to be more socially acceptable (Lavrakas, 2008; Zerbe and Paulhus, 1987). This behavior can be exhibited through behaviors observed while engaging in social tasks through self-deceptive and impression management styles, thus causing measurement error (Holtgraves, 2004; Paulhus, 2002). Social Desirability Bias (SDB) has been observed extensively in field and laboratory experiments. For example, Bateson et al. (2006) implemented a field experiment to evaluate how a pair of "watching" eyes affected donations to a university coffee fund. Subjects contribute to the coffee fund nearly three times as much when they are being "watched by the eyes" compared to the absence of the eyes watching control. Similarly, showing subjects computer-generated eyespots during dictator games increases generosity; players give twice as many tokens to their partners relative to the control group (Haley and Fessler, 2005; Sparks and Barclay, 2013). Controlling for SDB can be difficult. Therefore, it is important to understand the experimental context in which SDB may arise (Norwood and Lusk, 2011; Zizzo, 2010; Paulhus, 2002).

Researchers from other fields have found evidence of brief-lived social desirability bias with eye-tracking technology. In one study, only 36% of subjects who wore eye-tracking glasses looked at a provocative swimsuit calendar in a waiting room while 92% of those who did not wear the eye-tracking glasses looked at the calendar while being observed with a hidden camera (Risko and Kingstone, 2011). When the eye-tracker was turned off and the

subjects believed they were not being observed, 83% of the eye-tracking treatment looked at the calendar. However, even with wearable technology, this effect is short-lived; in a follow-up study, Nasiopoulos et al. (2015) found that after ten minutes without recalibration, a treatment group with wearable eye-trackers looked at the calendar at the same rate as those without eye-trackers. In another experiment, using eye-tracking equipment mounted on a computer screen, the knowledge of eye-tracking occurring during the experiment caused subjects to alter their gaze behavior to focus on female facial features rather than their chest and waist in the presence of eye-trackers even when the eye-trackers were not turned on (Wong and Stephen, 2019). In this experiment, the "eye-trackers" were made more salient by the researchers providing dual fake web-cameras with lights (either on or off depending on condition) on the top of the monitor while the actual unobtrusive eye-tracker secretly recorded eye-movements in all conditions. Both of these studies involved implicit deception and did not incentivize outcomes. Additionally, in both studies eye-trackers are more salient than they would be in a standard economics experiment in which the eye-tracker is generally a small black box at the base of a monitor (as shown for our experiment in Figure 1).

In economics, eye-tracking papers sometimes check their own experiments for Hawthorne effects induced by eye-tracking. For example, Wang et al. (2010) reports a null effect in a check comparing eye-tracking and non-eye-tracking participants in a sender-receiver game. Similarly, Harrison and Swarthout (2019) had the same group do their experiment with and without eye-tracking and found no difference in a risk lottery game. However, it is possible that there is publication bias for any individual experiment; papers not showing such an effect may be more likely to be published than those that do. Presently, to our knowledge, there is no economic study with the goal to comprehensively determine whether the physical presence and explicit knowledge of eye-tracking equipment influences subjects' decisions in incentivized laboratory settings. The economic games for this experiment are canonical games with varied *ex ante* expectations depending on the game. The games used in this experiment include the Dictator game, Ultimatum game, Public Goods game, Trust

game, Eckel and Grossman gambling risk task, Holt and Laury risk task, Double Auction, and a Cheating game. The null hypothesis is that there is no effect of eye-tracking on the economic behavior of participants for each economic game.

Overall, we do not find evidence of an eye-tracking effect in the economic behavior for seven of the eight games. Subjects neither behave more generously nor selfishly in the Dictator game, Ultimatum game, Public Goods game, and Trust game in the eye-tracking conditions. Meanwhile, they did not take less risk by choosing less risky gamble choices in the Eckel and Grossman gambling risk task, nor are they less likely to cheat by reporting a higher number in a Cheating game relative to the control conditions. In the Double Auction task, in the raw data, there are no statistical differences in the average profits, transaction price, bids or asks between the two groups, but there is a slight statistical difference in the transaction volume between the two groups. However, this difference disappears once the standard errors are clustered at the session label, and the indifference holds after controlling for the number of subjects and other demographic characteristics.

Only the Holt and Laury (HL) risk task showed a difference in means. Further exploration of these differences determined that both in the Holt and Laury risk task *and* the Eckel and Grossman risk task (using different sample populations), the number of calibration attempts was directly correlated with higher risk aversion. That is, for each additional calibration attempt, a participant's number of safe choices increased 0.2 points in both games. In the Holt and Laury game, these calibration outliers drove statistical differences between the treatment and control group. Further exploration of these results using choice process data on facial expressions collected by the facial engine coding, Affectiva AFFDEX, while participants were doing the study suggest that multiple calibration attempts are marginally correlated with feelings of fear, which are correlated with anxiety and guilt (Saxena et al., 2020), and that these feelings of fear are significantly correlated with taking less risky choices in *both* risk games.

In eye-tracking research, it is common practice to remove participants who do not have

eye-tracking accuracy above a certain "quality" threshold, such as 85% (according to iMotions). When we removed these observations from the treatment group, as would a researcher interested in the eye-tracking results themselves, the difference between the treatment and control groups in the Holt and Laury game is no longer significant. This result is likely because people who have difficulty calibrating also generate additional eye-tracking problems (Holmqvist et al., 2011; Hornof and Halverson, 2002; Nyström et al., 2013). Thus, for researchers who follow this standard practice of removing observations with poor eye-tracking, it appears that even with a risk assessment task, a Hawthorne effect caused by use of the eye-tracker is not a problem in eye-tracking research using these standard economics games.

The rest of the paper includes the following sections: Section 2 illustrates the experimental design, Section 3 describes the results of each game, Section 4 delves deeper into how calibration problems are correlated with risk aversion, and Section 6 concludes this paper.

2 Experimental Design

A total of 404 students from Texas A&M University were recruited to participate over the course of 50 sessions ranging from 4 to 16 participants.[1] Session times were available in the morning and afternoon during regular university hours of 8:00 A.M. – 5:00 P.M. Data were collected from July 2019 to February 2020. The experiment was conducted at the Human Behavior Lab at Texas A&M University. The lab has 16 stations equipped with computers mounted with eye-tracking devices and web cameras (Figure 1). Each station is surrounded by individual partitions six feet apart to prevent subjects from looking at other subjects. The experiment was computerized using Z-tree (Fischbacher, 2007).

The experiment consisted of eight economic games, including five social interaction games: Dictator, Double Auction, Public Goods, Trust and Ultimatum games; and three non-social

---

[1]Treatment Group 1 in the *eye-tracking condition* with 8 participants was dropped due to data loss by computer failure. So, we used 396 (248 in Group 1 and 148 in Group 2) and 49 sessions in total for our analysis. An *a priori* power analysis required a minimum of 164 subjects for two-player games (82 per role) and a minimum of 82 subjects for games other than two-player to have 80% power and a medium effect size (Cohen's D=0.50).

Figure 1: A picture of a station set up within the lab.

interaction games: Cheating game, and the Holt and Laury (HL) and Eckel and Grossman (EG) risk preference tasks. Conventional game instructions were used for all eight games (Andersen et al., 2018, 2011; Andreoni, 1995; Attanasi et al., 2016; Berg et al., 1995; Eckel and Grossman, 2002; Holt and Laury, 2002; Smith, 1962).[2] Because of the large number of games, we will present the details of game play for each game as well as predictions for social desirability bias along with its respective results in the results section.

We divided the eight games into two groups of four games to prevent subject fatigue. The games assigned to each group were selected to be balanced in terms of the game type and the amount of time required to complete the experiment. The games within Group 1 consisted of a one-shot Dictator game, Trust game, HL risk preference task, and a 10 period Double Auction. The games within Group 2 consist of the EG risk choice task, a 10 period Public Goods game, one-shot Ultimatum game, and 10 period Cheating game. The games for Group 1 were ordered from the most socially affecting to the least social affecting, while the games for Group 2 were ordered from the least socially affecting to the most social affecting.

---

[2]Experimental instructions are available in Appendix B.

The treatment conditions and Groups were randomized at the session level. The experiment was conducted using a between-subject design with an *Eye-tracking treatment condition* and a *no eye-tracking control*. For the *no eye-tracking condition*, the eye-tracking device and video camera were turned off before the subjects entered the lab. The subjects were read a script with no verbiage about eye-tracking equipment or calibration. To balance the two conditions, they looked at a blank screen for 4.5 minutes, the average amount of time it takes for calibration in the *eye-tracking condition*, to align the time required to complete the experiment. Table 1 presents the number of subjects for each experimental condition. Group 1 has 100 more subjects than Group 2.[3] However, the number of subjects between the treatment and control conditions are balanced.

Subjects were asked to sign one of two consent forms depending on their treatment assignment. The consent form for the *eye-tracking condition* included specific language consenting to the use of eye-tracking during the experiment. The *no eye-tracking condition* consent form had no language about eye-tracking procedures and equipment. Once the consent form was signed, subjects were randomly seated at one of the eye-tracking stations. For the *eye-tracking condition*, the subjects were read a script indicating that they would be calibrated before the session began. Before starting each game, the subjects were verbally reminded about the eye-tracking equipment and re-calibrated. We intentionally recalibrated after each game to remind the subjects of the presence of the equipment in the *eye-tracking condition* before each game, similar to Nasiopoulos et al. (2015) so that the presence of the equipment would be salient at the beginning of each game and so that any potential treatment effect would not diminish over the 90-minute study.

To make the average expected payoffs similar across games, the exchange rate for each game varied from the original game play. Once all the subjects finished making decisions for all games, the experimenter asked for a volunteer to draw a chip to determine the binding game and decision for payment. One of the four games was randomly selected for payment

---

[3]The Double Auction game in Group 1 requires larger number of subjects to play than do any of the games in Group 2.

then the subjects viewed their individual payoffs on their respective screens. After viewing their payoffs, the subjects were asked to complete a demographic survey. After all the subjects completed the survey, they received their payment in private. Subjects received a $10 show up fee plus the earnings they made based on their decisions of the binding game. Subjects were asked to sign a payment receipt form after they received their payment. The average payoff was $15.93 with a range of $14 to $20.

3   Game Results

Table 2 reports a test of whether the demographics, including gender, age in years, education, race, and income level, are balanced across the two conditions within group.[4] The *eye-tracking condition* in Group 1 has older, more educated, but less-White individuals than does the *no eye-tracking control*.[5] Additionally, the *eye-tracking condition* in Group 2 is younger and less educated than the *no eye-tracking control*. We control for these differences using regression analysis.

We present the results for each game in the following subsections based on *ex ante* expectations for each game and the order of play beginning with Group 1. Each game is examined with mean and distribution comparisons using Mann-Whitney test and Kolmogorov–Smirnov test. We also report OLS regression results for each game. Robust standard errors are clustered at the session level.

---

[4]Education is categorized as Freshman, Sophomore, Junior, Senior$^+$, Master, and Ph.D. Race includes White, African American, Asian, and Others. Income is categorized as <$45,000, $45,000-$49,000, $50,000-$59,000, and >$60,000.

[5]A balance test was conducted after collection in September. Age and race were unbalanced so data collection was initiated in October and again in February 2020. The balance check includes the results of a Kruskal-Wallis test of each Treatment Group of the difference for demographic characteristic. The detailed table is in Table 2.

3.1 Dictator game

The first game in Group 1 was the Dictator game, with instructions based on Andersen et al. (2018). In this game, subjects were matched randomly into pairs and were randomly assigned either the role of Player 1 or Player 2. Player 1 was asked to decide the number of tokens (out of 10) to split between themselves and Player 2. The exchange rate was 1 token equals $1. Both Players 1 and 2 read through the instructions then Player 1 made their decision while Player 2 was notified that Player 1 was making their choice. Subjects did not see the final result of the game unless it was chosen for payment at the end of the experiment. We hypothesized that if subjects feel the SDB in the *eye-tracking condition*, then they would transfer more tokens to Player 2 than in the *no eye-tracking control*. By doing this, they could be perceived as a generous person.

The giving behavior for Player 1 between the *eye-tracking* and *no eye-tracking conditions* is not significantly different. Figure 2 shows the comparison of the average amount of tokens sent between the conditions. There is no significant difference in the average number of tokens sent between the two conditions (MW, $p=0.37$; regression, $\beta = -0.325$, $p=0.66$). Like the mean comparison, we do not find differences in the distributions between the treatment and control (KS, $p=0.866$).
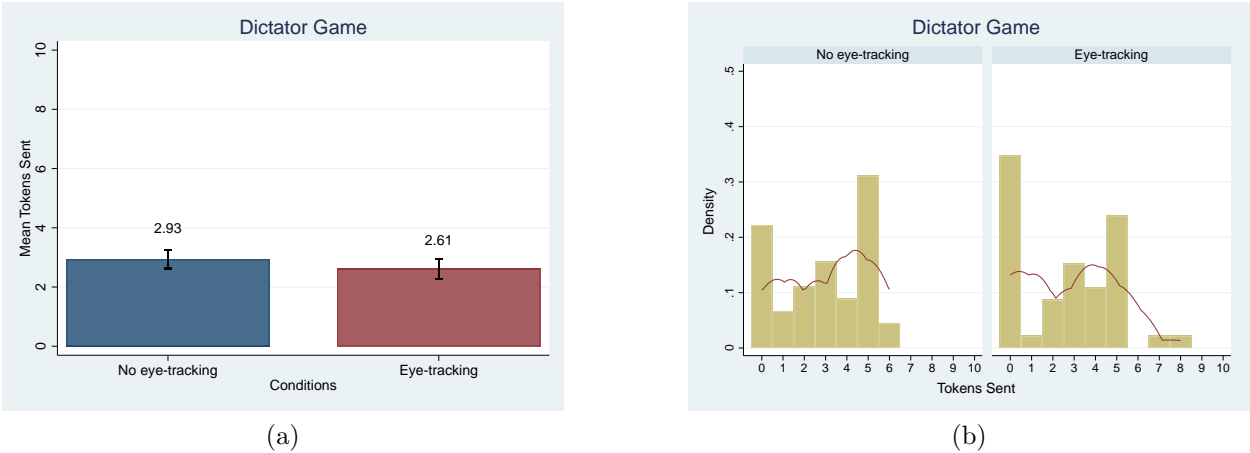

Figure 2: Mean and distribution comparisons in the number of tokens sent.

We also estimate OLS regressions with demographic controls in Table 3 and find no change in sign or significance from the regressions without controls clustered at the session level. Additionally, the average number of tokens sent in both conditions, 2.77 tokens, is within the range that has been previously found in the literature (Engel, 2011).

3.2 Trust game

Trust game instructions were based on Berg et al. (1995). Subjects were again paired with a random partner and randomly assigned either the role of Player 1 or Player 2. Both Players were endowed with 10 tokens. Player 1 was asked to decide the number of tokens they want to transfer to Player 2. Then, Player 2 received tripled the number of tokens from Player 1. Player 2 selected the number of tokens to return to Player 1 from the available funds they had. The exchange rate was 1 token equals $0.5. In this game, there is a chance that subjects might behave in a more socially desirable way in the *eye-tracking condition*. Player 1 in the *eye-tracking condition* might act more trusting or generous by sending a higher number of tokens to Player 2 than Player 1 in the *no eye-tracking control* and Player 2 may return more tokens to Player 1 in the *eye-tracking condition* compared to the *no eye-tracking control*.

We do not find evidence of differences in behavior for either Players 1 or 2 across conditions. As a measure of trust, we used the fraction of tokens they could send for Player 1. As a measure of trustworthiness for Player 2, we used the fraction of what they could return. Figure 3 presents the mean difference in the proportions of tokens sent and returned between conditions. There is no difference in the proportion of tokens sent by Player 1 (MW, $p=0.76$) and no difference in the proportion of tokens returned by Player 2 across conditions (MW, $p=0.93$). Along with the mean comparisons, the distributions are not different between the *eye-tracking condition* and *no eye-tracking control* for either Player 1 (KS, $p=0.35$; regression, $\beta = -0.37$, $t=0.77$) or Player 2 (KS, $p=1.00$; regression, $\beta = -0.013$, $t=0.27$).

Results from OLS regressions controlling for different demographics, as provided in columns (2) and (3) of Table 3, are similarly insignificant for both Player 1 and Player
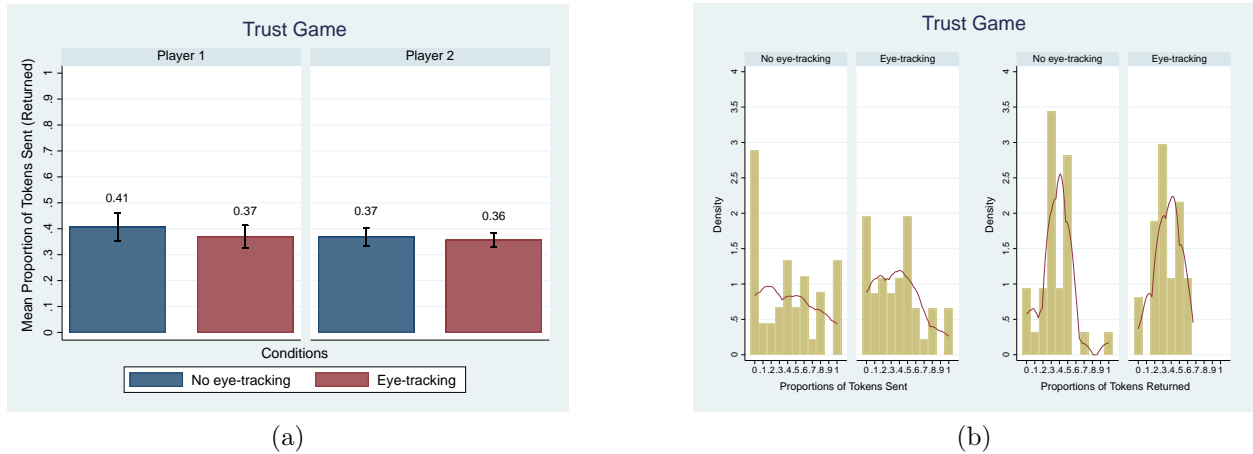
Figure 3: Mean and distribution comparisons in the proportions of tokens sent and returned.
Note. For reciprocity, those who received 0 tokens in Player 2 treated as missing values (Burks, Carpenter and Verhoogen, 2003). The x-axis in the histogram represents the proportion of tokens the players received.

2 although the sign flips for Player 2. Eye-tracking does not affect behavior in the Trust game regardless of the inclusion of control variables.[6] In addition, the average proportion of tokens sent and returned in our study is consistent with previous studies (Eckel and Wilson, 2004; Johnson and Mislin, 2011).[7]

3.3   Holt and Laury Risk Task

Subjects played the risk lottery game from Holt and Laury (2002) (HL). This risk task consists of 10 Decisions with two safe and risky options labeled as A and B. As subjects progress through the Decisions, the probabilities for the payoffs in Option A and B change. Subjects are asked to choose one of the options in each Decision. Unlike the conventional HL game, in which all ten choices are displayed on a single screen, the game was modified so

---

[6]There is also no significant effect on trustworthiness in any regression results.

[7]Eckel and Wilson (2004) report that Player 1 sent an average of 70 percent of the endowment and the reciprocity, including only those who returned at least 10 tokens (endowment for Player 2), of Player 2 is 45 percent with the Texas A&M University students. On the other hand, the meta-analysis report that the Player 1 sent 0.5 proportion of the endowment and the Player 2 returned 0.37 proportion of the available funds they could return, which is only the tripled amount the Player 2 received (Johnson and Mislin, 2011). In our study, for the Player 1, the average number of tokens sent in our study, 0.39, is lower than but similar to the meta-analysis. For the Player 2, the average proportion of tokens returned, accounting the tripled amount they received, is 0.36, which aligns with the meta-analysis.

that each lottery decision was displayed on separate screens to ensure incentive compatibility (Brown and Healy, 2018). The exchange rate was 1 token for $4. The subjects read through the instructions and went through one example. Then they participated in 10 real rounds in sequential order with a separate screen for each decision, meaning that the number of safe choices could vary between 0 and 10. We did not have strong predictions about the direction of socially desirable behavior; a person might prefer to choose more safe options in the *eye-tracking condition* than in the control in order to be perceived as a less risk-taking person. However, it is also possible that they might prefer to be perceived as more risk-taking and thus choose the opposite.

In contrast to the previous games, we observe an eye-tracking effect in HL. On average, subjects in both treatments showed risk averse behavior, meaning the average number of safe choices is greater than 4 (Holt and Laury, 2002), as can be seen in Figure 4a. However, subjects in the *eye-tracking condition* have more risk averse behavior than those in the *no eye-tracking condition* (MW, $p=0.04$; $\beta=0.50$, $t=2.18$). This difference also appears in the frequency distribution in Figure 4b (KS, $p=0.07$). The distribution is slightly skewed to the right in the *no eye-tracking condition* compared to the *eye-tracking condition*.
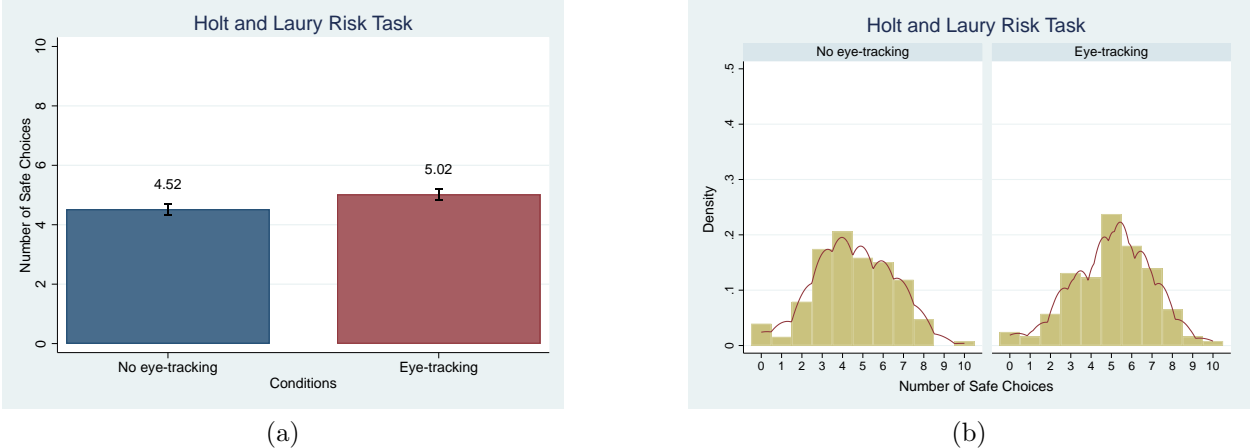

Figure 4: Mean and distribution comparisons in the number of safe choices.

This difference between the treatment and control groups remains when demographic characteristics are controlled for ($\beta=0.545$, $t=2.31$), as shown in column (4) of Table 3. This

difference still holds marginally for estimations that exclude participants showing inconsistent behavior, aka "multiple switchers" (β=0.505, t=1.85 without controls; β=0.490, t=1.73 with demographic controls).[8] In our study, the average number of safe choices for both conditions is 4.76. This number is similar to the original risk task study (Holt and Laury, 2002).[9] However, subjects make more safe choices in the *eye-tracking condition* when there are multiple failed calibration attempts. We further investigate the number of failed calibrations attempts taking eye-tracking and emotions data into account in Section 4.

3.4   Double Auction

The last game in Group 1 was the Double Auction, with instructions based on Smith (1962) and Attanasi et al. (2016). In this game, subjects were randomly assigned the role of either Buyer or Seller. The market consisted of 10 periods, each 2 minutes in length, and each subject had 1 unit of a fictitious good to buy or sell in each period depending on their role. The value for the Buyer varied from 1 to 8 and the cost for the Seller varied from 3 to 10. Buyers could not buy the goods at a price higher than their values and Sellers could not sell their goods at a price lower than their costs. Participants were informed that they could post their bids and asks in increments of 0.50 tokens. Before each period started, subjects were informed of their role and value or cost; these were fixed over the 10 periods and were displayed on the left side of the screen. Each period subjects could freely and simultaneously place a bid or ask in the market by inputting their bid or ask price at the bottom of the screen. If there was a price they wanted to buy or sell at, they could click the "Sell at this price" or "Buy at this price" button on the bottom of the screen. All transactions in the period were displayed on the right side of the screen. The remaining time was displayed on

---

[8]A multiple switcher is someone who made multiple switches between option A and option B in which they demonstrate inconsistent behavior. An example of a multiple switcher would be someone who chose option A until Decision 4 then switched to option B in Decision 5 and switched back to option A in Decision 6 and so on. It is common to report results with and without multiple switchers included (Holt and Laury, 2002; Kassas et al., 2019).

[9]The number of safe choices in Holt and Laury (2002) was 5.2 which is lower than the 6.3 safe choices found with Texas A&M University students in Kassas et al. (2019).

the top of the right side of the screen. The exchange rate was 1 token equals \$1. Based on standard economic theory, we would expect no difference in the competitive equilibrium for the Double Auction game based on social desirability bias.

We use the average over the 10-periods of each of five dimensions to test for an eye-tracking effect in the Double Auction: profits, transaction price, transaction volume, bids, and asks. First, we examine the average profits over the 10 periods. We do not find any treatment effect. Figure 5 shows no differences in the mean and distribution of profits between conditions (MW, $p=0.54$ and KS, $p=0.39$). There is also no eye-tracking effect on the average transaction price (MW, $p=0.713$ and KS, $p=0.965$), average asks (MW, $p=0.862$ and KS, $p=0.973$), or average bids (MW, $p=0.326$ and KS, $p=0.612$). See the figures for these additional variables in Figures A1-A4 in the appendix A.
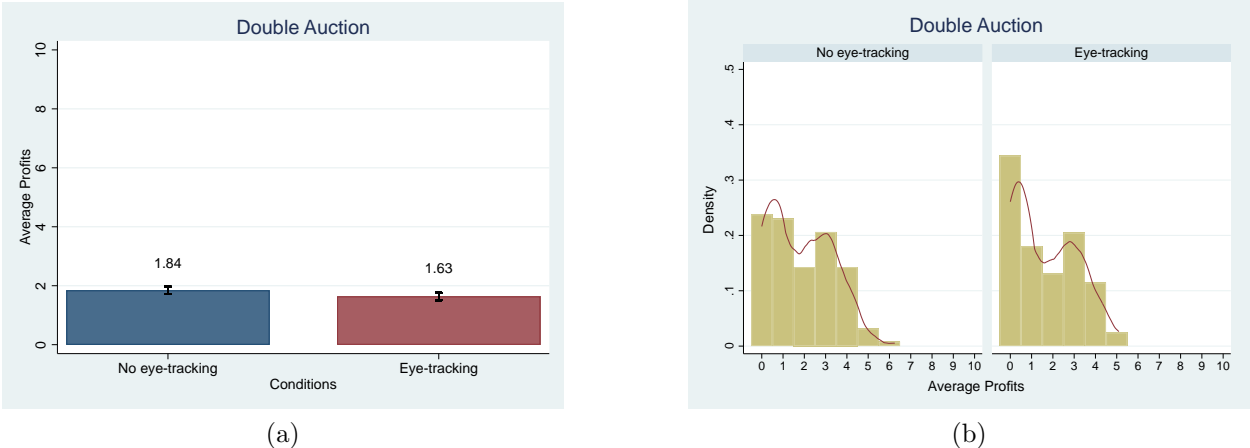
Figure 5: Mean and distribution comparisons in the average profits of 10 periods.

We find marginally significant differences in the mean transaction volume between the *eye-tracking* and *no eye-tracking conditions* (MW, $p=0.073$). However, once the standard errors are clustered at the session level using OLS, these effects are no longer statistically significant ($\beta = -0.064$, $t= -1.43$ for volumes), nor are they significant when we control for the number of subjects ($\beta = -0.031$, $t= -1.17$).[10] All auction results remain insignificant

---

[10]The number of participants in a market should mechanically increase transaction volume, so to the extent that market sizes differ between the treatment and control groups, it should be controlled for when transaction volume is an outcome (Wang and Yau, 2000).

when demographic controls are added.[11]

3.5   Eckel and Grossman Gambling Task

The first game in Group 2 was the Eckel and Grossman Gambling Task (EG), based on Eckel and Grossman (2002). Like the HL Risk Task, this task was a gamble choice task. Subjects were presented with six different gamble choices with two different payments in each choice and a 50-50 chance of each occurring. The alternatives increased in risk and expected return from Gamble 1 to Gamble 5. Gamble 6 had the same expected return as Gamble 5, but higher risk. Subjects were asked to choose only one gamble choice. This method of eliciting subject risk preferences has the advantage of requiring minimal math skills (Reynaud and Couture, 2012; Eckel et al., 2012; Eckel and Grossman, 2008, 2002). The choice set in our investigation contained no loss of money (or tokens) as opposed to "loss" framework found in previous literature; the gambles were displayed as circles increasing in expected payoffs clockwise instead of in a table format. The exchange rate was 1 token equals $0.50. As with HL, subjects might pick more risk averse choices in the *eye-tracking condition* than subjects in the *no eye-tracking condition* or vice versa if they wish to be perceived as more or less risk seeking.

For EG, unlike HL, subjects' behavior is not significantly different in the *eye-tracking condition* compared to the *no eye-tracking condition*. The average mean gamble choice is measured to examine the treatment effect. Like the HL Risk Task, subjects reveal risk averse behavior by choosing Gamble Choices 1 and 2, as shown in Figure 6, which has been reverse coded (subtracted from 7) to show the mean safe gamble choice in order to make it comparable to HL. Again, unlike HL, there are no significant differences in the mean gamble choices or the distributions between conditions (MW, $p=0.56$ and KS, $p=0.38$).

Similarly, including demographic controls in an OLS framework provides no evidence

---

[11]We also analyzed the convergence to the equilibrium price and quantity along with the demand and supply. The results look as expected for double auctions, with no obvious differences between treatment and control groups and are available from the authors.

of an eye-tracking effect, as demonstrated in column (6) of Table 3. The most frequently chosen gamble choice, 2 (reverse coded as 5 in our charts), and the average gamble choice, 2.53 (reverse coded as 3.47 in our charts), are close to that from previous studies, including one using the same sample population (Kassas et al., 2019; Eckel and Grossman, 2002).

(a)

(b)
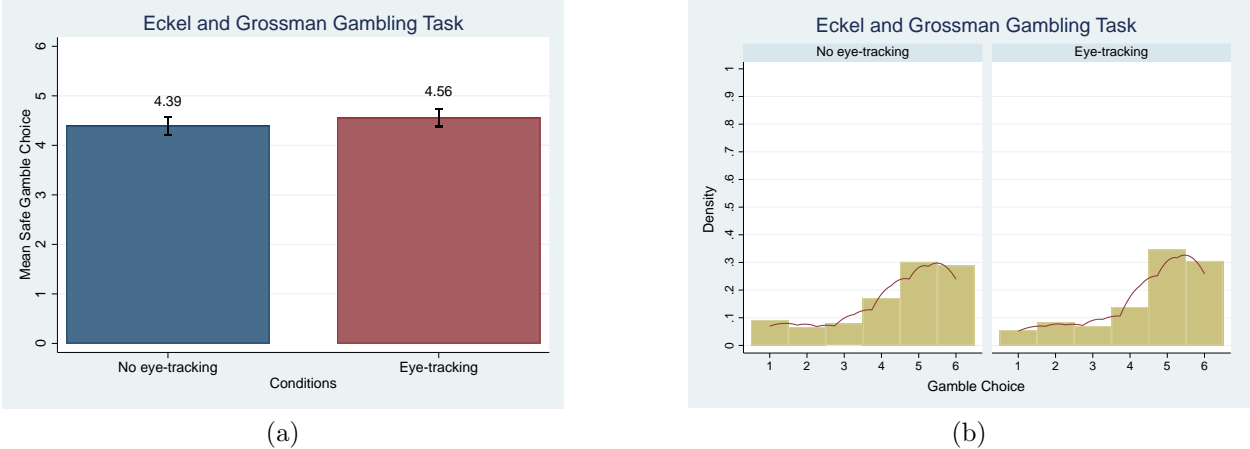
Figure 6: Mean and distribution comparisons in the gamble choices.

3.6 Public Goods game

The instructions for the Public Goods game were based on Andreoni (1995). Subjects were randomly assigned into groups of 4 each round and they played the game for 12 rounds (2 practice and 10 real rounds). Each period each subject was endowed with 100 tokens. They were informed that they had two accounts, public and private, and were asked to allocate the number of tokens they wanted to go to each account. All tokens in the private account yielded a return of 10 cents. Each token invested into the public account, by all members, had a yield of half a cent. At the end of each round, the participants viewed the number of tokens invested into their private account, the public account, and their earnings for that period. The rounds were not timed. The exchange rate was 1 token for $0.10. If there is social desirability bias, we would expect subjects to contribute a larger number of tokens to the public account in the *eye-tracking condition* compared to the *no eye-tracking condition.*

Figure 7: Mean and distribution comparisons in the average number of tokens kept in 10 periods.

Subjects in the *eye-tracking condition* do not behave differently than subjects in the *no eye-tracking condition*. The 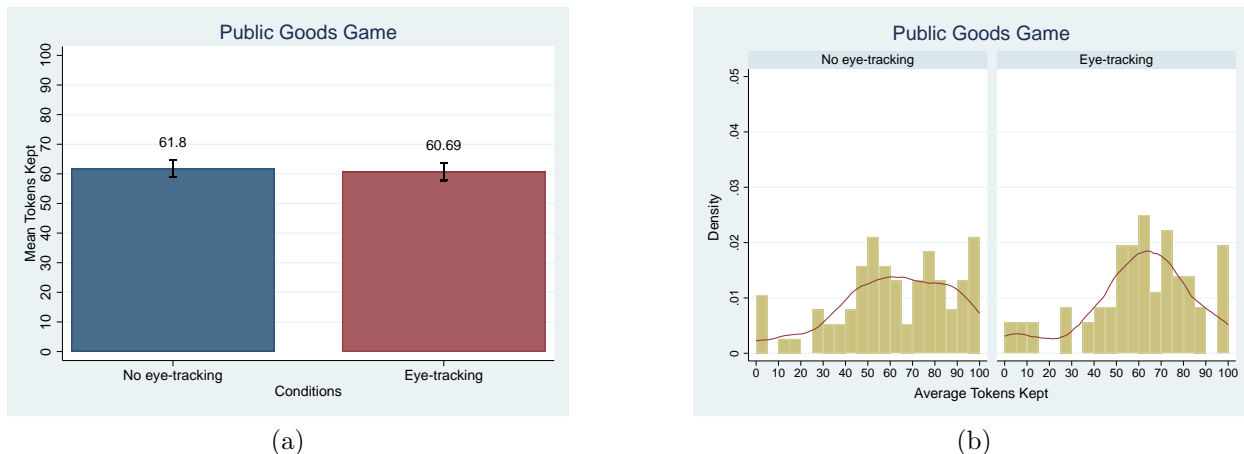mean number of tokens kept in the private account is the metric used in this game. Figure 7 presents that no difference exists in the mean tokens kept and the distribution of the tokens kept (MW, $p=0.99$ and KS, $p=0.79$). According to the distributions, allocating all tokens to the private account was chosen the most in both conditions. Keeping 50 tokens was the second most frequent choice. The distribution of tokens kept is centered around 65 in the *eye-tracking condition* whereas the tokens kept in the range from 30 to 100 is more evenly distributed for the *no eye-tracking condition*. Controlling for demographic characteristics does not change the lack of treatment effect, as shown in column (7), Table 3. The percentage of tokens kept in our study, 61.26%, is higher than a previous meta-analysis on the Public Goods game, but it is similar to studies conducted at Texas A&M University (Zelmer, 2003; Eckel et al., 2015).[12]

3.7   Ultimatum game

Game instructions for the Ultimatum game were modified from Andersen et al. (2011). Subjects were randomly matched with a partner and were randomly assigned to be either

---

[12]On average, 37.7% of the endowment was contributed in Zelmer (2003), whereas on average 67% was contributed among the Texas A&M University students in Eckel et al. (2015).

Player 1 or Player 2. Similar to the Dictator game, Player 1 was endowed with 10 tokens and asked to select the number of tokens to transfer to Player 2. However, in this game, Player 2 could either reject or accept the offer from Player 1. If Player 2 rejected the offer, then both Players would receive nothing, otherwise the allocation was made according to the amount proposed by Player 1. The exchange rate was 1 token for $1. Social desirability bias would suggest that Player 1 would send more tokens in the *eye-tracking condition* compared to the *no eye-tracking condition*. The prediction for Player 2 is ambiguous.
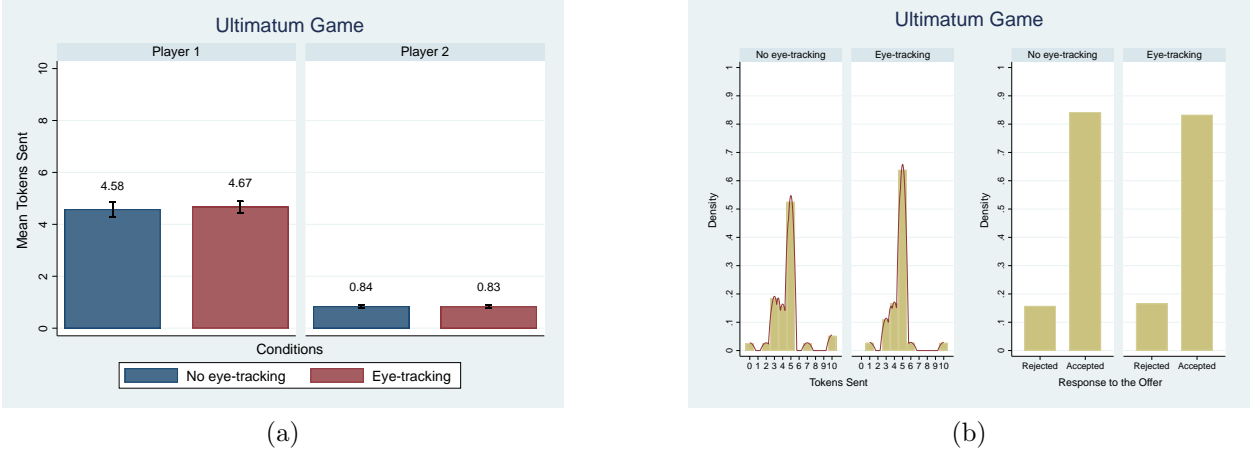
Figure 8: Mean and distribution comparisons in the number of tokens sent and the acceptance of the offer.

Note. For player 2, acceptance is coded as 1 and rejection is coded as 0.

Subjects again did not behave differently between the eye-tracking treatment and control in the Ultimatum game. We measured the mean number of tokens Player 1 sent and the acceptance rate of Player 2. Figure 8 shows the mean and distribution comparisons in the number of tokens sent and the acceptance rate between conditions. Both the average number of tokens sent by Player 1 and the acceptance rate of Player 2 are no different between conditions (MW, $p=0.44$ for Player 1 and MW, $p=0.81$ for Player 2). Similarly, the distributions are not different between conditions for both Player 1 and Player 2 (KS, $p=0.99$ for Player 1 and KS, $p=1.00$ for Player 2). OLS results available in columns (8) and (9) of Table 3 and probit results (available upon request) controlling for demographic

18

characteristics do not find evidence of an eye-tracking effect. The average endowment in the offers, 46.2%, and acceptance rates, 14%, in our study align with a previous meta-analysis of the Ultimatum game (Oosterbeek et al., 2004).[13]

3.8   Cheating game

The final game in Group 2 was the Cheating game with instructions adapted from Aksoy and Palma (2019), which is based on Fischbacher and Föllmi-Heusi (2013). This game involves 10 periods. In each period, subjects see a random sequence of numbers consisting of 0, 2, 4, 6, and 8. They are asked to report the first number they see in each period and are told that the number that they report is their payment. Since the order of the numbers is randomly generated, and there is no enforcement or punishment for misreporting the number that participants actually saw first, there is a chance they can lie by reporting a higher number than what they actually saw and profit from it. As promised, subjects were paid for the value that they reported. The exchange rate was 1 token equals $1. SDB would suggest that subjects would be less likely to lie, that is, the average reported number would be lower, in the *no eye-tracking condition* than in the *eye-tracking condition.*

In our results, eye-tracking does not affect the subjects' propensity to lie. The average number of tokens reported in 10 periods is used to measure the treatment effect in this game. The expected average number of tokens from the random number placement is 4 tokens (i.e. the expected value of the five numbers that appear on the computer screen in random order: 0, 2, 4, 6, 8). In our study, the average number of tokens reported in both conditions for the 10 periods is higher than the expected number of tokens, about 4.75. Figure 9 shows the mean and distribution comparisons between conditions with the expected likelihood of 20% for each number under truth telling. There is no difference in the average number of tokens reported between conditions (MW, $p=0.70$). The distributions for both conditions are slightly skewed to the right and there is no difference in the distribution (KS, $p=0.98$).

---

[13]The average of 40 percent of the endowment is offered by Player 1 and the average of 16 percent of the offers are rejected in Oosterbeek et al. (2004).

Again, controlling for demographic characteristics does not change the results, as shown in column (10) of Table 3. We also confirm that the average number of tokens reported in our study aligns with previous studies showing that there is a tendency for lying, but the magnitude is small (Abeler et al., 2019).

(a)   (b)
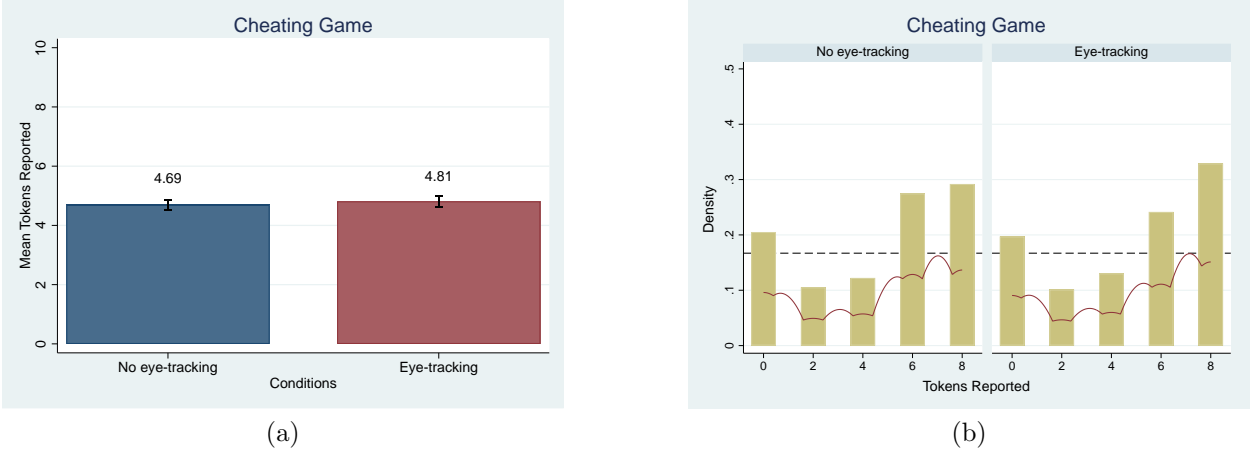
Figure 9: Mean in the average number of tokens reported in 10 periods and distribution comparisons in the number of tokens reported in 10 periods.
Note. Dashed line indicates the expected density of each token occurring, in which a truthful distribution (0.167) (Abeler et al., 2019; Aksoy and Palma, 2019).

4   Number of Calibration Attempts and Risk Aversion

In the previous section, we show that there are no eye-tracking effects in the games except for the HL Risk Task. In this section, we dive deeper into the HL result and explore the effect of the number of calibration attempts on risk aversion in both the HL and EG risk tasks. We first show that the number of failed calibration attempts is correlated with higher measured risk aversion in both games. We then examine potential selection bias into who has difficulty eye-tracking and do not find significant demographic differences between those who do and do not have difficulty calibrating in our sample. We next remove participants who have trouble during the calibration stage of the experiment and find that the significant eye-tracking effect in HL goes away. These results are robust to different thresholds for

defining calibration difficulty as explained in section 4.3. Then, given that people who have trouble calibrating on the eye-tracker often do not have usable eye-tracking data and are thus excluded from reported results, we drop likely excluded participants, those with poor eye-tracking, from our analyses to see if the significant difference in HL goes away, which it does. Finally, we use unobtrusive facial expression analysis software, which does not require calibration, to explore why increased eye-tracking calibration attempts may affect measured risk aversion.

4.1 Poor calibration is correlated with higher measured risk aversion

We first show that the number of failed calibration attempts is correlated with higher measured risk aversion in both risk games. We calculate the number of failed calibration attempts for HL and EG by summing the failed calibration attempts prior to the start of each game. Using OLS estimation in Table 4, columns (1)-(6), we find that the number of failed calibration attempts significantly increases the number of safe choices in both risk games by about 0.2, even with the inclusion of the demographic characteristics and the removal of participants with inconsistent preferences. Recall that HL and EG were played by different participants and in different orders, with HL as the third game in Group 1, and EG as the first game in Group 2.

4.2 Demographic characteristics of poor eye-tracking calibrators

We examine the characteristics of subjects across the number of failed calibration attempts for HL and EG to determine if differences in risk aversion are driven by differences in demographic characteristics related to calibration difficulty rather than being caused by failed calibration attempts themselves. In the literature, calibration difficulties are caused by things like unusually wet eyes, highly reflective glasses, air bubbles under contact lenses,

droopy eyelids, small pupils, or lighting conditions (Harezlak et al., 2014; Nyström et al., 2013; Holmqvist et al., 2011). Using the Kruskal-Wallis test for means comparison of the cumulative number of calibrations, we find no significant differences across demographic characteristics of gender, age, educational status, race, or income in the cumulative number of calibrations for either HL or EG. These results are available from the authors.[14] There may be demographic characteristics that we did not observe that are correlated with both difficulty calibrating and increased risk aversion, but we do not find any correlations for these first-order possibilities.

4.3   Results of removing subjects with poor eye-tracking/calibration data

Given that the number of failed calibration attempts is significantly related to the risk averse behavior in both HL and the EG and does not seem to be correlated with major demographic characteristics that would cause selection bias, we explore the potential solution of simply removing poor calibrators directly and then removing those with poor eye-tracking data overall, something that is standard in eye-tracking research.

The cumulative number of calibration attempts ranges between 1 and 10 for HL, the only game in which an eye-tracking effect was found. Using the Mann-Whitney test and OLS estimation, where standard errors are clustered at session level, we find that the eye-tracking effect disappears when we exclude subjects who had a cumulative number of calibration attempts greater than 6 by HL (~ 10 percent of observations), with and without demographic controls included (MW, $p=0.130$). Table 5, columns (1)-(4) present the OLS results gradually removing people with the most difficulty calibrating (results, not shown, are similar including

---

[14]We also examined the average number of calibration attempts for each game. The average number of calibration attempts per subject per game varies from 1 to 6. Using the Kruskal-Wallis test for means comparison of the average number of calibrations, we find no significant differences across demographic characteristics in the average number of calibrations. Testing the demographic characteristics separately, the average number of calibrations is marginally lower among African Americans when compared to Whites in the OLS estimation result. These results are also available from the authors.

controls). Just removing the one person with 10 calibration attempts does not remove the significant eye-tracking result (β=0.501, t=2.17), but removing subjects with 9 and 8 calibration attempts reduces the magnitude of the effect and makes it marginally significant (β=0.441, t=1.88, and β=0.439, t=1.83, respectively). Removing subjects with 7 calibration attempts reduces it further to β=0.383, t=1.55. This pattern continues (in results not shown) with the sign of the eye-tracking effect eventually flipping negative when people with more than 2 calibration attempts are removed (N = 138).

Multiple calibrations were attempted when the eye-tracker could not perfectly catch the subject's eyes. Thus, multiple calibration attempts are correlated with low quality eye-tracking data, something that is generally measured as the fraction of frames the eyes were detected (valid samples) over the total number of recorded frames (total samples) (Tobii user's manual; Tobii Pro, 2018; Karch 2018; Nyström et al. 2013).[15] Dropping low quality data (i.e. lower than 85 percent threshold according to iMotions) is standard practice in many eye-tracking studies. We first test for an eye-tracking effect in HL with this standard 85% threshold cut, which drops 18 percent of the observations. As shown in Table 5, column (6), the eye-tracking effect goes away when these low eye-tracking data quality are removed (β=0.084, t=0.44). For researchers who want a lower quality threshold, we also tested the effect of removing observations at the 80% threshold shown in Table 5 column (5), and found that although the magnitude is larger (β=0.383, t=1.55), it is not significant and is similar to what is found by dropping observations with 7 or more calibration attempts.[16] The results presented in Table 5 columns (5) and (6) assume a normal filter, but results with medium and high filters are similarly insignificant, with magnitudes diminishing as expected, and are available from authors. We also test the EG data to make sure that the null result is robust

---

[15]https://www.tobiipro.com/siteassets/tobii-pro/product-descriptions/tobii-pro-tx300-product-description.pdf, https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-studio-user-manual.pdf, Tobii Pro, 2018: https://www.tobiipro.com/siteassets/tobii-pro/accuracy-and-precision-tests/tobii-pro-spectrum-accuracy-and-precision-test-report.pdf/?v=1.1

[16]The Pearson correlation between calibration attempts greater than 6 and the low eye tracking data quality is 0.055 (p=0.023), validating that the subjects who experienced difficulties during the calibration stage have the low eye tracking data quality.

to dropping calibrators or poor quality data and it is. Similarly, null eye-tracking effect in the other 8 games are also robust to removal of poor calibrators and those with poor quality eye-tracking data (results available from authors).

4.4 Biometric analysis for subjects who had trouble with eye-tracking calibration

Having shown that an eye-tracking effect is correlated with difficulty calibrating with our two groups in two separate risk games, and that removing these poor calibrators removes the eye-tracking effect, we further investigate plausible mechanisms underlying the cumulative calibration attempts and decision-making behaviors in the two risk games. In section 4.2, we showed no relationship between the simple demographics we measured and difficulty calibrating. In this section, we turn to the potential effects of emotions utilizing measures of facial expression.[17]

We use Affectiva AFFDEX, the facial expression engine analysis, to measure the emotional state of subjects. We examine these metrics on the first instruction screen presented right after the calibration stage. We test, as our outcome variables, seven measures of emotional states (joy, anger, surprise, fear, contempt, disgust, and sadness). In Table 6 Panels 1 and 2, we show a series of simple bivariate regressions with difficulty calibrating (defined as being in the worst 10% of calibrators for each game) as the X variable and our seven different measures of emotional states as Y variables in each column. Panel I shows results for HL, while Panel II shows results for EG.

Subjects who had trouble calibrating show fear with marginal significance for both games. Panel I, Column (4) shows that subjects who have difficulty calibrating show a marginally significant increase of 2.93 in the Fear measure (t=1.74) for HL, more than 2x the average fear level of 1.25 for HL, while Panel II, Column (4) shows a marginally significant increase

---

[17]We also collected and tested eye tracking pupil data to measure arousal and engagement and find mixed results depending on game, but given that difficulty with calibration is correlated with inability to track pupil data, we do not want to put weight on these results.

of 1.84 in the same Fear measure (t=1.86), which is also more than 2x the average fear level of 0.80 for EG. At the same time, we do not observe any significant differences in Joy, Anger, Surprise, Contempt, Disgust, or Sadness for HL, and while Joy is significant in EG in Panel II, Column (1), the value is much larger than the not significant corresponding value for HL. While it is true that it is likely in any set of seven regressions that one result may be marginally significant, we think that it is suggestive that the same emotional state is marginally significant in both risk games with the same sign and similar magnitudes.

There are several possible paths that could lead this fear measure to increase risk aversion. First, psychology literature recognizes secondary emotions as the combination of these measured primary emotions. In this regard the coexistence of joy and fear in the EG task could provide an indication of guilt (Coulter and Pinto, 1995), and this guilt from difficulty calibrating could be driving results. Second, Saxena et al. (2020) find that anxiety is an auxiliary emotional state that is strongly correlated (0.97 correlation) with fear. The strong correlation between fear and anxiety may be a possible explanation for the link between fear and higher levels of risk aversion (Eisenbach and Schmalz, 2016). Finally, if trouble calibrating leads directly to fear (possibly of not getting paid), and fear then leads to more risk averse behavior, then these findings align with earlier research showing that negative emotions drive people to choose a greater number of safe choices (Campos-Vazquez and Cuilty, 2014; Bassi et al., 2013; Kuhnen and Knutson, 2011). Still, we caution that these emotional results are at best suggestive and merit future research.

5   Discussion and Conclusion

Our study reveals that eye-tracking equipment does not affect individual behavior in economic decision making for seven out of eight popular economic games. Participants, on average, behaved the same in both *eye-tracking* and *no eye-tracking conditions*, particularly after implementing standard procedures, such as controlling for the number of participants

per session or clustering at the session level in the Double Auction. Additionally, results from the individual games are in line with previously done studies using the same games. In one of our risk preference games, HL, there is a significant difference between the outcomes of the *eye-tracking* and *no eye-tracking* groups. Further exploration into this outcome suggests that subjects who have difficulty with the eye-tracking calibration procedure are driving these differences. The differences in the number of safe choices are still prevalent when we control for demographic differences between subjects who have difficulty calibrating and those who do not, suggesting that the calibration effect is not just driven by selection on observables. Instead, potentially increased fear (or anxiety/guilt) from having multiple calibration trials results in the risk averse behavior in HL.

We began our research with a hypothesis about SDB, and that any eye-tracking effect, if it exists, will be more prevalent in games that are more susceptible to SDB. We feel confident suggesting that SDB emerging from eye-tracking is not a concern in standard economics games. Even the cheating game, where we expected the highest amount of SDB, did not show differences between *eye-tracking* and *no eye-tracking* groups. We speculate, like Andersen et al. (2011), that the equipment in these types of eye-tracking experiments might be inconspicuous enough that it is not intrusive to the subjects as the equipment rests on the bottom of the computer monitor and is not physically placed on the subjects' bodies. Additionally, university students are of a generation of individuals who have been continuously immersed in technology from a young age. They might not be as averse to being observed through technology because of the normalized presence of webcams on their phones and computers as well as security cameras in stores and homes.

Although we find no evidence of SDB with eye-tracking, we do find evidence that calibration difficulty can potentially create negative emotions that can affect the risk preferences of bad calibrators. The one game that showed significant differences between the *eye-tracking* group and *no eye-tracking* group was a risk aversion game, HL. Although testing the effect of calibration on risk aversion was not part of our pre-analysis plan, this correlation holds

across two different games (HL and EG) taken by two different sets of people (Group 1 and Group 2), lending credence to the idea that this is a real effect. We suspect that differences in significance between the two games are likely because of the different order that the games were presented, with one of them presented as the first game and the other showing the cumulative effects of calibration difficulties, though we cannot rule out differences in the games themselves or differences between the two subject pools.

Our results suggest that in games that are affected by fear, difficulty with calibration may affect the results of risk preference instruments. However, given that calibration difficulty is correlated with eye-tracking data quality, this effect is limited and easy to remedy. It appears that the standard process of removing poor-quality eye-tracking data also removes subjects likely to be affected by calibration problems. Another potential solution to this problem is that researchers could keep track of the number of calibration attempts per subject, and should test their results for robustness with and without observations with abnormal numbers of calibration attempts or controlling for the number of calibration attempts.[18] Finally, as eye-tracking software and hardware get better, this calibration effect should diminish.

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond, "Preferences for Truth-Telling," *Econometrica*, 2019, *87* (4), 1115–1153.

Aksoy, Billur and Marco A Palma, "The Effects of Scarcity on Cheating and In-Group Favoritism," *Journal of Economic Behavior & Organization*, 2019, *165*, 100–117.

Andersen, Steffen, Seda Ertaç, Uri Gneezy, Moshe Hoffman, and John A List, "Stakes Matter in Ultimatum Games," *American Economic Review*, 2011, *101* (7), 3427–39.

_ , Uri Gneezy, Agne Kajackaite, and Julie Marx, "Allowing for Reflection Time Does Not Change Behavior in Dictator and Cheating Games," *Journal of Economic Behavior and Organization*, 2018, *145*, 24–33.

Andreoni, James, "Cooperation in Public-Goods Experiments: Kindness or Confusion?," *The American Economic Review*, 1995, pp. 891–904.

---

[18]We obviously cannot control for calibration attempts in the *eye-tracking* vs. *no eye-tracking* results here given that the *no eye-tracking* group does not calibrate at all. However, most researchers using eye-tracking will not have a *no eye-tracking* group and thus should be able to include those controls.

Attanasi, Giuseppe, Samuele Centorrino, and Ivan Moscati, "Over-the-Counter Markets vs. Double Auctions: A Comparative Experimental Study," *Journal of Behavioral and Experimental Economics*, 2016, *63*, 22–35.

Bassi, Anna, Riccardo Colacito, and Paolo Fulghieri, "'O Sole Mio: An experimental Analysis of Weather and Risk Attitudes in Financial Decisions," *The Review of Financial Studies*, 2013, *26* (7), 1824–1852.

Bateson, Melissa, Daniel Nettle, and Gilbert Roberts, "Cues of Being Watched Enhance Cooperation in a Real-World Setting," *Biology Letters*, 2006, *2* (3), 412–414.

Berg, Joyce, John Dickhaut, and Kevin McCabe, "Trust, Reciprocity, and Social History," *Games and Economic Behavior*, 1995, *10* (1), 122–142.

Brown, Alexander L. and Paul J. Healy, "Separated Decisions," *European Economic Review*, 2018, *101*, 20–34.

Campos-Vazquez, Raymundo M and Emilio Cuilty, "The Role of Emotions on Risk Aversion: a Prospect Theory Experiment," *Journal of Behavioral and Experimental Economics*, 2014, *50*, 1–9.

Chen, Chun Ting, Chen Ying Huang, and Joseph Tao Yi Wang, "A Window of Cognition: Eyetracking the Reasoning Process in Spatial Beauty Contest Games," *Games and Economic Behavior*, 2018, *111*, 143–158.

Chen, Wei James and Ian Krajbich, "Computational Modeling of Epiphany Learning," *Proceedings of the National Academy of Sciences of the United States of America*, 2017, *114* (18), 4637–4642.

Coulter, Robin Higie and Mary Beth Pinto, "Guilt Appeals in Advertising: What Are Their Effects?," *Journal of Applied Psychology*, 1995, *80* (6), 697–705.

Devetag, Giovanna, Sibilla Di Guida, and Luca Polonio, "An Eye-Tracking Study of Feature-Based Choice in One-Shot Games," *Experimental Economics*, 2016, *19* (1), 177–201.

Eckel, Catherine C. and Philip J. Grossman, "Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk," *Evolution and Human Behavior*, 2002, *23* (4), 281–295.

_ **and** _ , "Men, Women and Risk Aversion: Experimental Evidence," *Handbook of Experimental Economics Results*, 2008, *1* (C), 1061–1073.

Eckel, Catherine C and Rick K Wilson, "Is Trust a Risky Decision?," *Journal of Economic Behavior & Organization*, 2004, *55* (4), 447–465.

Eckel, Catherine C., Haley Harwell, and Castillo G. José Gabriel, "Four Classic Public Goods Experiments: A replication Study," *Research in Experimental Economics*, 2015, *18*, 13–40.

_ , **Philip J. Grossman, Cathleen A. Johnson, Angela C.M. de Oliveira, Christian Rojas, and Rick K. Wilson**, "School Environment and Risk Preferences: Experimental Evidence," *Journal of Risk and Uncertainty*, 2012, *45* (3), 265–292.

Eisenbach, Thomas M. and Martin C. Schmalz, "Anxiety in the Face of Risk," *Journal of Financial Economics*, 2016, *121* (2), 414–426.

Engel, Christoph, "Dictator Games: A Meta Study," *Experimental Economics*, 2011, *14* (4), 583–610.

Fengs, Gary, "Eye Tracking: A brief Guide for Developmental Researchers," *Journal of Cognition and Development*, 2011, *12* (1), 1–11.

Fiedler, Susann, Andreas Glöckner, Andreas Nicklisch, and Stephan Dickert, "Social Value Orientation and Information Search in Social Dilemmas: An Eye-Tracking Analysis," *Organizational Behavior and Human Decision Processes*, 2013, *120* (2), 272–284.

Fischbacher, Urs, "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments," *Experimental Economics*, 2007, *10* (2), 171–178.

___ **and Franziska Föllmi-Heusi**, "Lies in disguise—an experimental study on cheating," *Journal of the European Economic Association*, 2013, *11* (3), 525–547.

Haley, Kevin J and Daniel MT Fessler, "Nobody's watching?: Subtle cues affect generosity in an anonymous economic game," *Evolution and Human behavior*, 2005, *26* (3), 245–256.

Harezlak, Katarzyna, Pawel Kasprowski, and Mateusz Stasch, "Towards accurate eye tracker calibration–methods and procedures," *Procedia Computer Science*, 2014, *35*, 1073–1081.

Harrison, Glenn W. and J. Todd Swarthout, "Eye-Tracking and Economic Theories of Choice Under Risk," *Journal of the Economic Science Association*, 2019, *5* (1), 26–37.

Holmqvist, Kenneth, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer, *Eye Tracking: A Comprehensive Guide to Methods and Measures*, OUP Oxford, 2011.

Holt, Charles A and Susan K Laury, "Risk Aversion and Incentive Effects," *The American Economic Review*, 2002, *92* (5), 1644–1655.

Holtgraves, Thomas, "Social Desirability and Self-Reports: Testing Models of Socially Desirable Responding," *Personality and Social Psychology Bulletin*, 2004, *30* (2), 161–172.

Hornof, Anthony J and Tim Halverson, "Cleaning up systematic error in eye-tracking data by using required fixation locations," *Behavior Research Methods, Instruments, & Computers*, 2002, *34* (4), 592–604.

Hristova, Evgenia and Maurice Grinberg, "Information Acquisition in the Iterated Prisoner's Dilemma Game: An Eye-Tracking Study," in "Proceedings of the 27th Annual Conference of the Cognitive Science Society" Lawrence Erlbaum Hillsdale, NJ 2005, pp. 983–988.

Hu, Yingyao, Yutaka Kayaba, and Matthew Shum, "Nonparametric Learning Rules from Bandit Experiments: The Eyes Have It!," *Games and Economic Behavior*, 2013, *81* (1), 215–231.

Johnson, Noel D. and Alexandra A. Mislin, "Trust Games: A Meta-Analysis," *Journal of Economic Psychology*, 2011, *32* (5), 865–889.

Kaminska, Olena and Tom Foulsham, "Understanding Sources of Social Desirability Bias in Different Modes: Evidence from Eye-Tracking," Technical Report, ISER Working Paper Series 2013.

Karch, Jessica M, "Beyond Gaze Data: Pupillometry as an Additional Data Source in Eye Tracking," in "Eye Tracking for the Chemistry Education Researcher," ACS Publications, 2018, pp. 145–163.

Kassas, Bachir, Marco A Palma, and Maria Porter, "Happy to Take Some Risk: Investigating the Dependence of Risk Preferences on Mood Using Biometric Data," 2019.

Knoepfle, Daniel T, Joseph Tao yi Wang, and Colin F Camerer, "Studying Learning

in Games Using Eye-Tracking," *Journal of the European Economic Association*, 2009, *7* (2-3), 388–398.

Król, Michał and Magdalena Ewa Król, "A Novel Approach to Studying Strategic Decisions with Eye-Tracking and Machine Learning," *Judgment and Decision Making*, 2017, *12* (6), 596–609.

Kuhnen, Camelia M and Brian Knutson, "The Influence of Affect on Beliefs, Preferences, and Financial Decisions," *Journal of Financial and Quantitative Analysis*, 2011, *46* (3), 605–626.

Lahey, Joanna N. and Douglas Oxley, "The Power of Eye Tracking in Economics Experiments," *American Economic Review*, 2016, *106* (5), 309–313.

Lavrakas, Paul J, *Encyclopedia of Survey Research Methods*, Sage Publications, 2008.

Lejarraga, Tomás, Michael Schulte-Mecklenbeck, and Daniel Smedema, "The PyeTribe: Simultaneous Eyetracking for Economic Games," *Behavior research methods*, 2017, *49* (5), 1769–1779.

Nasiopoulos, Eleni, Evan F Risko, Tom Foulsham, and Alan Kingstone, "Wearable computing: Will it make people prosocial?," *British Journal of Psychology*, 2015, *106* (2), 209–216.

Norwood, F Bailey and Jayson L Lusk, "Social Desirability Bias in Real, Hypothetical, and Inferred Valuation Experiments," *American Journal of Agricultural Economics*, 2011, *93* (2), 528–534.

Nyström, Marcus, Richard Andersson, Kenneth Holmqvist, and Joost van de Weijer, "The Influence of Calibration Method and Eye Physiology on Eyetracking Data Quality," *Behavior Research Methods*, 2013, *45* (1), 272–288.

Oosterbeek, Hessel, Randolph Sloof, and Gijs Van De Kuilen, "Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis," *Experimental Economics*, 2004, *7* (2), 171–188.

Paulhus, Delroy L, "Socially desirable responding: The evolution of a construct," *The role of constructs in psychological and educational measurement*, 2002, *49459.*

Peshkovskaya, Anastasia G, Tatiana S Babkina, Mikhail G Myagkov, Ivan A Kulikov, Ksenia V Ekshova, and Kyle Harriff, "The Socialization Effect on Decision Making in the Prisoner's Dilemma Game: An Eye-Tracking Study," *PloS One*, 2017, *12* (4), e0175492.

Pieters, Rik and Michel Wedel, "Goal Control of Attention to Advertising: The Yarbus Implication," *Journal of Consumer Research*, 2007, *34* (2), 224–233.

Polonio, Luca, Sibilla Di Guida, and Giorgio Coricelli, "Strategic Sophistication and Attention in Games: An Eye-Tracking Study," *Games and Economic Behavior*, 2015, *94*, 80–96.

Popovic, Nathalie, "Belief Elicitation and Information Search in Normal-Form Games. An Eye-Tracking Study." PhD dissertation 2014.

Reutskaja, Elena, Rosemarie Nagel, Colin F. Camerer, and Antonio Rangel, "Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study," *American Economic Review*, 2011, *101* (2), 900–926.

Reynaud, Arnaud and Stéphane Couture, "Stability of Risk Preference Measures: Results from a Field Experiment on French Farmers," *Theory and Decision*, 2012, *73* (2), 203–221.

Risko, Evan F. and Alan Kingstone, "Eyes Wide Shut: Implied Social Presence, Eye Tracking and Attention," *Attention, Perception, and Psychophysics*, 2011, *73* (2), 291–296.

Rosch, Jonathan L. and Jennifer J. Vogel-Walcutt, "A Review of Eye-Tracking Applications as Tools for Training," *Cognition, Technology and Work*, 2013, *15* (3), 313–327.

Saxena, Piyush, Sarthak Dabas, Devansh Saxena, Nithin Ramachandran, and Sheikh Iqbal Ahamed, "Reconstructing Compound Affective States using Physiological Sensor Data," *Proceedings - 2020 IEEE 44th Annual Computers, Software, and Applications Conference, COMPSAC 2020*, 2020, pp. 1241–1249.

Sharafi, Zohreh, Zéphyrin Soh, and Yann-Gaël Guéhéneuc, "A Systematic Literature Review on the Usage of Eye-Tracking in Software Engineering," *Information and Software Technology*, 2015, *67*, 79–107.

Sickmann, Jörn and Huynh Bich Ngan Le, "Eye-Tracking in Behavioural Economics and Finance-A Literature Review," 2016.

Smith, Vernon L, "An Experimental Study of Competitive Market Behavior," *Journal of Political Economy*, 1962, *70* (2), 111–137.

Sparks, Adam and Pat Barclay, "Eye Images Increase Generosity, but Not for Long: The Limited Effect of a False Cue," *Evolution and Human Behavior*, 2013, *34* (5), 317–322.

Stewart, Neil, Simon Gächter, Takao Noguchi, and Timothy L Mullett, "Eye Movements in Strategic Choice," *Journal of Behavioral Decision Making*, 2016, *29* (2-3), 137–156.

Tanida, Shigehito and Toshio Yamagishi, "Testing Social Preferences Through Differential Attention to Own and Partner's Payoff in a Prisoner's Dilemma Game," *Letters on Evolutionary Behavioral Science*, 2010, *1* (2), 31–34.

Wang, George HK and Jot Yau, "Trading Volume, Bid–Ask Spread, and Price Volatility in Futures Markets," *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 2000, *20* (10), 943–970.

Wang, Joseph Tao Yi, Michael Spezio, and Colin F. Camerer, "Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games," *American Economic Review*, 2010, *100* (3), 984–1007.

Wedel, Michel, "Attention Research in Marketing: A Review of Eye Tracking Studies," *SSRN Electronic Journal*, 2014, pp. 1–28.

Wong, Hoo Keat and Ian D. Stephen, "Eye Tracker as an Implied Social Presence: Awareness of Being Eye-Tracked Induces Social-Norm-Based Looking Behaviour," *Journal of Eye Movement Research*, 2019, *12* (2).

Zelmer, Jennifer, "Linear Public Goods Experiments: A Meta-Analysis," *Experimental Economics*, 2003, *6* (3), 299–310.

Zerbe, Wilfred J and Delroy L Paulhus, "Socially Desirable Responding in Organizational Behavior: A Reconception," *The Academy of Management Review*, 1987, *12* (2), 250–264.

Zizzo, Daniel John, "Experimenter Demand Effects in Economic Experiments," *Experimental Economics*, 2010, *13* (1), 75–98.

Tables

Table 1: Experimental sequence for Groups 1 and 2 including the number of sessions and observations between treatment conditions.

Conditions Calibration	**Group 1**		**Group 2**	
	Treatment (Calibration)	Control (4.5 min. wait)	Treatment (Calibration)	Control (4.5 min. wait)
	Dictator		Eckel and Grossman Gambling	
	Trust		Public Goods	
	Holt and Laury Risk		Ultimatum	
	Double Auction		Cheating	
Number of Sessions	14[A](122)	18[B](126)	8 (72)	9 (76)

Note. Observations are in parentheses. For Group 1, we have collected additional sessions, running only the HL Risk Task and Double Auction to reach the power.
[A] We collected 11(92).
[B] We collected 13(90) for Dictator and Trust game.

Table 2: Demographic Balance Test Across Treatments.

	Group 1			Group 2		
	No Eye tracking	Eye tracking	Kruskal-Wallis test	No Eye tracking	Eye tracking	Kruskal-Wallis test
Male[A]	0.452	0.426	P= 0.722	0.387	0.521	P= 0.161
Age (years)	21.452	22.098	P= 0.017	23.868	21.861	P= 0.029
Education			P= 0.013			P= 0.038
Freshman	0.206	0.172		0.158	0.222	
Sophomore	0.222	0.074		0.053	0.097	
Junior	0.167	0.148		0.079	0.139	
Senior[+]	0.167	0.221		0.197	0.208	
Master	0.119	0.32		0.355	0.222	
Ph.D.	0.119	0.066		0.158	0.111	
Race			P= 0.010			P= 0.336
White	0.54	0.369		0.342	0.431	
Black	0.056	0.066		0.013	0.028	
Asian	0.31	0.402		0.526	0.431	
Others	0.095	0.164		0.118	0.111	
Income[B]			P= 0.372			P= 0.118
<$45,000	0.389	0.375		0.473	0.375	
$45,000-$49,000	0.048	0.117		0.122	0.042	
$50,000-$59,000	0.063	0.125		0.054	0.083	
>$60,000	0.5	0.383		0.351	0.5	
N	126	122		76	72	

Note. Means and P-values from Kruskal-Wallis test are reported. Senior[+] includes both senior and 5th year in undergraduate.
[A] Group 2 had 75 observations for males in the no eye-tracking condition and 71 observations for males in the eye-tracking condition.
[B] Group 1 had 120 observations for income level in the eye-tracking condition. Group 2 had 74 observations for income level in the no eye-tracking condition.

Table 3: Regression estimates for all games including demographic controls.

Variables	Coefficients (S.E.)									
	(1) Dictator game	(2) Trust game (Player 1)	(3) Trust game (Player 2)	(4) Holt and Laury	(5) Double Auction Profits	(6) Eckel and Grossman	(7) Public Goods	(8) Ultimatum game (Player 1)	(9) Ultimatum game (Player 2)	(10) Cheating game
Eye tracking	-0.254	-0.053	0.01	0.545**	-0.165	0.125	3.283	0.475	-0.275	0.136
	(0.528)	(0.074)	(0.062)	(0.236)	(0.149)	(0.254)	(4.590)	(0.489)	(0.436)	(0.280)
Male	-0.363	0.051	0.018	-0.693**	-0.137	-0.348	-12.515**	0.354	-0.373	-0.388
	(0.414)	(0.066)	(0.038)	(0.292)	(0.223)	(0.269)	(4.577)	(0.372)	(0.382)	(0.272)
Age	0.143	-0.014	-0.010*	0.097*	-0.053	-0.065	1.347*	-0.012	-0.086**	-0.031
	(0.131)	(0.014)	(0.005)	(0.049)	(0.044)	(0.040)	(0.670)	(0.064)	(0.035)	(0.029)
Sophomore	0.000	-0.175	0.011	-0.642	0.518	-0.332	5.079	0.836		-0.475
	(0.802)	(0.124)	(0.073)	(0.431)	(0.340)	(0.355)	(11.129)	(0.723)		(0.428)
Junior	-1.270*	-0.113	-0.068	-0.566	0.326	-0.87	-4.562	-0.236	-0.315	-0.098
	(0.713)	(0.134)	(0.094)	(0.468)	(0.336)	(0.554)	(7.777)	(0.521)	(1.116)	(0.374)
Senior+	-1.846**	-0.077	-0.036	-0.674	0.251	-0.734***	-3.098	0.768	-0.151	0.492
	(0.701)	(0.143)	(0.081)	(0.413)	(0.395)	(0.243)	(9.099)	(0.701)	(0.730)	(0.317)
Master	-1.826	0.046	-0.012	-0.858	0.568	-0.307	-7.937	1.448**	0.638	-0.217
	(1.395)	(0.204)	(0.090)	(0.520)	(0.417)	(0.351)	(9.611)	(0.681)	(0.742)	(0.399)
Ph.D.	-1.709	-0.021	0.015	-0.917	1.053	0.219	-19.318	0.908	0.622	0.335
	(1.718)	(0.258)	(0.114)	(0.682)	(0.770)	(0.648)	(11.569)	(0.701)	(0.615)	(0.620)
Black	2.321*	-0.021	-0.140**	-0.572	-1.029***	-1.396**	4.158	-0.818		0.429
	(1.227)	(0.134)	(0.066)	(0.690)	(0.287)	(0.616)	(8.131)	(0.546)		(0.499)
Asian	-0.787	-0.209**	0.016	-0.162	0.023	0.339	8.736	0.448	0.564	0.775**
	(0.680)	(0.079)	(0.068)	(0.354)	(0.257)	(0.276)	(6.085)	(0.442)	(0.550)	(0.341)
Others	0.342	-0.082	-0.039	-0.365	-0.236	-0.483	0.221	-0.297		0.389
	(0.541)	(0.079)	(0.079)	(0.539)	(0.329)	(0.320)	(10.474)	(0.397)		(0.427)
$45,000-49000	0.341	0.338**	0.196**	-0.431	-0.204	-0.75	-2.436	1.028*	-1.214	-0.326
	(0.892)	(0.125)	(0.084)	(0.541)	(0.441)	(0.693)	(4.417)	(0.539)	(0.803)	(0.255)
$50,000-59000	0.37	0.059	-0.11	0.238	0.143	0.286	4.85	0.277		-0.386
	(0.813)	(0.138)	(0.086)	(0.351)	(0.456)	(0.587)	(9.676)	(0.600)		(0.518)
>$60,000	-0.371	0.044	-0.031	0.142	-0.190	0.14	-8.978	0.97	0.682	0.231
	(0.591)	(0.144)	(0.061)	(0.354)	(0.282)	(0.265)	(7.578)	(0.832)	(0.424)	(0.338)
Constant	1.339	0.789**	0.608***	3.352***	2.807***	6.328***	38.494**	3.089	2.486**	5.051***
	(2.635)	(0.327)	(0.104)	(0.974)	(0.928)	(0.723)	(14.659)	(2.173)	(1.041)	(0.632)
N	91	91	68	246	246	144	144	71	59	144
R-squared	0.205	0.211	0.205	0.08	0.063	0.147	0.156	0.222	0.163	0.104

Note. Coefficients from OLS estimations are reported. Standard errors are presented in parentheses and clustered at the session level. The proportion of the number of tokens sent is used for Player 1, whereas the reciprocity is used for Player 2 in Trust Game. The probit model is used for Player 2 in Ultimatum Game. Pseudo R-squared is reported in column (9). Senior+ includes both senior and 5th year in undergraduate. * Statistically significant at 10% level; ** at 5% level; *** at 1% level.

Table 4: Relationship between number of safe choices and cumulative number of calibration attempts prior to game.

	(1)	(2)	(3)	(4)	(5)	(6)
		Holt and Laury			Eckel and Grossman	
	All Sample		No Switchers			
# calibration attempts	0.185***	0.182**	0.192**	0.201**	0.220**	0.226*
	(0.057)	(0.063)	(0.068)	(0.069)	(0.085)	(0.106)
Controls	N	Y	N	Y	N	Y
N	122	120	99	97	72	71
R-squared	0.04	0.19	0.04	0.16	0.05	0.26

Note. Coefficients from OLS estimations for the treatment groups only are reported. Standard errors are presented in parentheses and clustered at the session level. No switchers removes subjects with inconsistent preferences. Controls include male, age, year in school indicators, race indicators, and categorical family income variables. * Statistically significant at 10% level; ** at 5% level; *** at 1% level.

Table 5: Holt and Laury number of safe choices OLS with universe restrictions.

|	(1)	(2)	(3)	(4)	(5)	(6)
	Cutting by Calibration Attempts				Cutting by data quality %	
	#Cal≥10	#Cal≥9	#Cal≥8	#Cal≥7	80%	85%
Eye-Tracking	0.501**	0.441*	0.439*	0.383	0.320	0.084
	(0.230)	(0.234)	(0.240)	(0.247)	(0.211)	(0.189)
N	247	243	238	235	205	196
R-squared	0.02	0.01	0.01	0.01	0.01	0.00

Note. Coefficients from OLS estimations are reported. Standard errors are presented in parentheses and clustered at the session level. Columns (5) and (6) use a normal filter which discards the observations that have both eyes have values of 4 or 2 (i.e. (4,4) or (2,2)). Results using a medium and high filter are similarly insignificant, though the magnitudes using the high filter (both eyes have values of 0) are smaller. * Statistically significant at 10% level; ** at 5% level; *** at 1% level.

Table 6: Effect of difficulty calibrating on emotions as measured by face recognition software.

	(1) Joy	(2) Anger	(3) Surprise	(4) Fear	(5) Contempt	(6) Disgust	(7) Sadness
	Panel I: Holt and Laury						
Trouble Calibrating	2.091	-0.762	0.688	2.930*	-1.832	-0.743	-1.605
	(2.530)	(1.779)	(2.337)	(1.687)	(2.475)	(0.867)	(3.016)
	Panel II: Eckel and Grossman						
Trouble Calibrating	7.423***	-0.275	-2.662	1.840*	-1.911	-0.083	-0.203
	(1.888)	(0.570)	(3.037)	(0.988)	(2.755)	(0.646)	(0.297)

Note. Each column represents the results from a separate OLS regression. Robust standard errors are in parentheses. Trouble calibrating is 1 if the subject is in the worst 10% of calibrators for each game. Number of observations are 101 in the first panel and 62 in the second panel. Observation numbers decrease compared to Table 1 because of lack of sensitivity in the facial recognition software. * Statistically significant at 10% level; ** at 5% level; *** at 1% level.
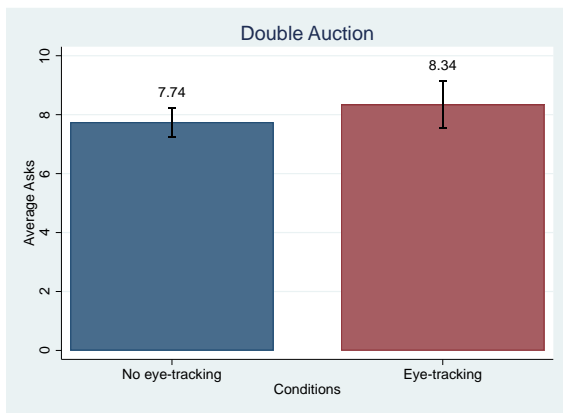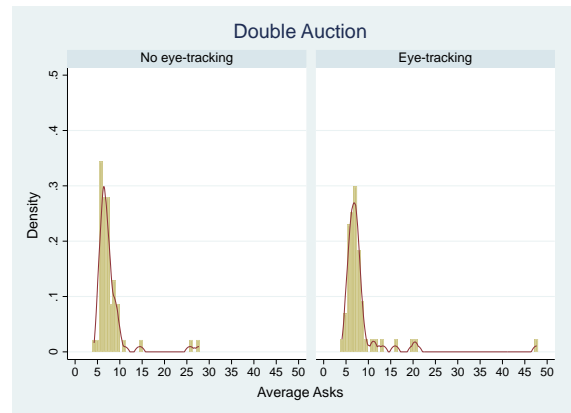
Appendix A.
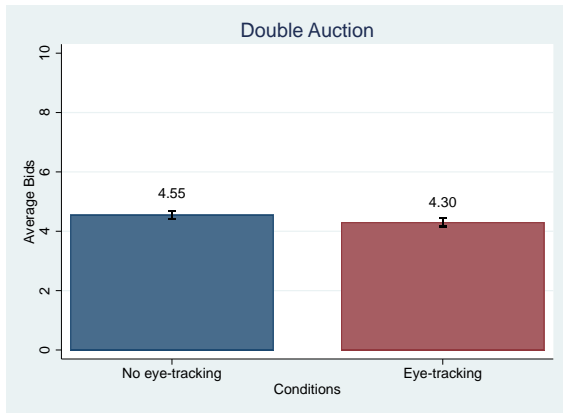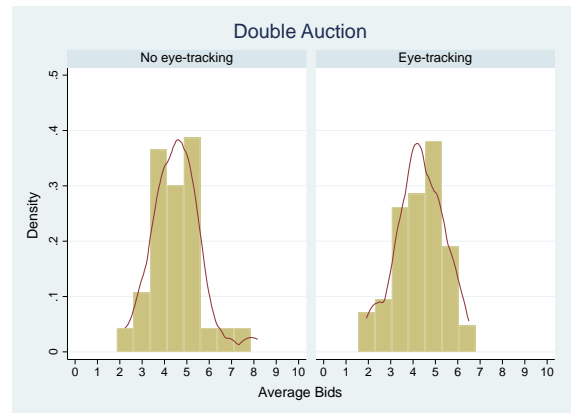
Figure A1: Mean and distribution comparisons in bids.

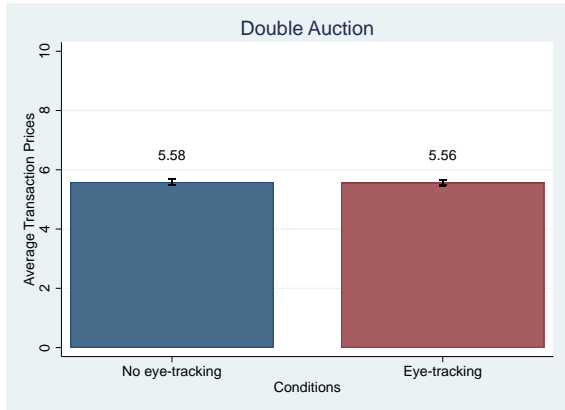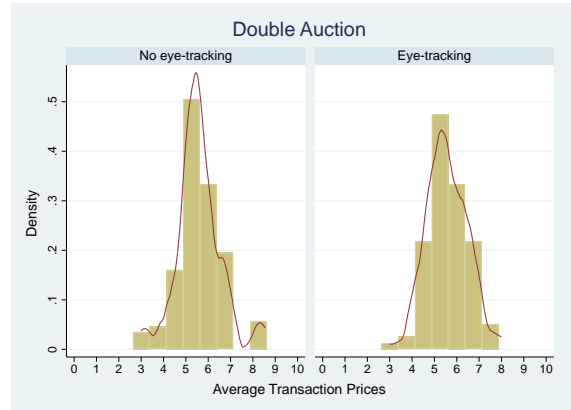
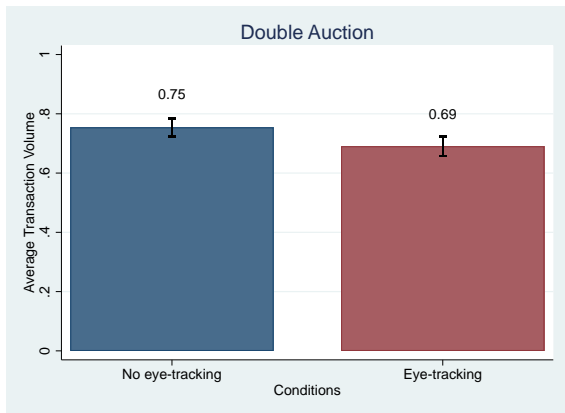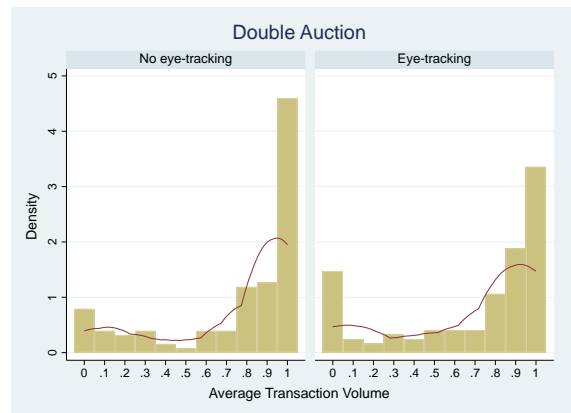Figure A2: Mean and distribution comparisons in asks.

(a)

(b)

Figure A3: Mean and distribution comparisons in transaction prices.

(a)

(b)

Figure A4: Mean and distribution comparisons in transaction volume.

Appendix B. Abbreviated Instructions

This section presents a transcription of the abbreviated instructions for each game. The full script is available from the authors upon request.

General Instructions

Welcome and thank you for participating in our study. You will receive $10 for participating; this will be yours to keep. You will also have the opportunity to make more earnings based on your decisions, the decisions of the other players, and luck. So, please pay attention to the instructions.

Today you will play 4 games. At the end of the session, one of the games will be randomly selected for payment. The total amount you earn today will be paid to you in cash and privately at the end of the experiment. Please do not talk to other participants.

1. Dictator game

[*Page 1*] **Instructions**
In this game there are two players: Player 1 and Player 2. You will be randomly matched with another person in this room. You will not be told the identity of the person that you are matched with and the other person will not know your identity. One of you will be randomly assigned the role of Player 1 and the other the role of Player 2. Player 1 begins with 10 tokens and Player 2 begins with 0 tokens. There will be one period only.

In this game, 1 token is equal to $1.

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

[*Page 2*]
Player 1's Decision:
Player 1 will be asked to decide to split 10 tokens between Player 1 and Player 2. Player 1 can choose any integer amount between 0 and 10 tokens.
Player 2's Decision:
Player 2 will be told the offer and has no choice to make.
Payoffs:
That is, if this game is selected for payment, Player 1's payoff (not including show-up fee) = 10 tokens minus the number of tokens transferred to Player 2. Player 2's payoff (not including show-up fee) = the amount Player 1 transferred.
Please raise your hand if you have any questions, otherwise please click NEXT to continue.

[*Page 3*] Once all decisions have been made the game will end. We will calculate payoffs based on the decisions made. Again, you will be randomly paid for one of the games at the conclusion of the experiment.

[*Page 4*] You are Player 1./You are Player 2.

[*Page 5*]

| | Player 1's screen | | | | | | | | | | Player 2' screen |
You are **Player 1**.
Now, please select the number of tokens you want to send to Player 2.

Number of tokens to transfer to Player 2										
0	1	2	3	4	5	6	7	8	9	10

Player 1 is making a decision.

[*Page 6*] This concludes Game 1. Please wait for the experimenter to come.

2. Trust game

[*Page 1*] **Instructions**
In this game there are two players: Player 1 and Player 2. You will be randomly matched with another person in this room. You will not be told the identity of the person that you are matched with and the other person will not know your identity. One of you will be randomly assigned the role of Player 1 and the other the role of Player 2. Your payoffs will be determined by the decisions that you both make. Player 1 and Player 2 both begin with 10 tokens.

In this game, 2 tokens are equal to $1.

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

[*Page 2*] **Player 1's Decision:**
Player 1 moves first. Player 1 may send some, all, or none of the 10 tokens to Player 2. Each token sent to Player 2 will be tripled.For example, if Player 1 sends 2 tokens, Player 2 receives 6 tokens = (2 tokens × 3). If Player 1 sends 9 tokens, Player 2 receives 27 tokens = (9 tokens × 3).Player 2 will then decide how many tokens to send back to Player 1 and how many tokens to keep. Player 1 indicates how much to send to Player 2 by typing the appropriate amount of tokens on the decision screen.

[*Page 3*] **Player 2's Decision:**
Player 2 begins with 10 tokens. In addition, Player 2 receives three times the amount sent by Player 1. Player 2 may send back some, all, or none of the tripled amount to Player 1 (Player 2 keeps the 10 tokens they started with). Any tokens sent back to Player 1 will not be tripled. Once Player 1 has sent some, none, or all of their tokens, Player 2 will then decide how many tokens to send back to Player 1 and how many tokens to keep.

[*Page 4*] Once all decisions have been made the game will end. We will calculate payoffs based on the decisions made. Again, you will be randomly paid for one of the games at the conclusion of the experiment.

[*Page 5*] You are Player 1./You are Player 2.

[*Page 6. Player 1's screen*]
You are **Player 1.**
You can send some, all, or none of the 10 tokens. Please indicate how many tokens to send to Player 2 in the form below. You can send any integer amount between 0 and 10 tokens.
  Enter the amount between 0 and 10. ☐
Please raise your hand if you have any questions, otherwise please click NEXT to continue.

You are **Player 2.**

Player 1 sent [ ] tokens, thus you received [ ] tokens.

How much do you want to send back to Player 1?

  Enter the amount between 0 and 10. ☐

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

This concludes Game 2. Please wait for the experimenter to come.

3. Holt and Laury Risk Task

Instructions

In this game you will make ten decisions. Please make your choices carefully. The Decision Screen shows a choice between Option A and Option B. You will make one choice and record it in the choice column. Then you will move to the next Decision Screen. Please raise your hand if you have any questions, otherwise please click NEXT to continue.

Even though you will make ten decisions, if this game is selected for payment only one decision will affect your earnings. One of the decisions you make will be selected randomly for your payment. Since you do not know in advance which decision will be selected, your best approach is to make each decision as if it is the one that will be selected for payment. Each decision has an equal chance of being selected.

In this game, 1 token is equal to $4.

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

If this game is selected for payment, we will ask one of you to volunteer to draw a chip from a bag. The bag contains 10 chips numbered 1 through 10, one for each decision in this game. The chip will determine which decision will be the real decision. After pulling the chip, your payment will be determined based on your decision and a roll of a dice. This will be explained in the following screens.

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

Once all decisions have been made the game will end. We will calculate payoffs based on the decisions made. Again, you will be randomly paid for one of the games at the conclusion of the experiment.

The following gives you an example of the decision screen. It reproduces the first decision. Option A pays either 2.00 toknes or 1.60 tokens. It depends on the roll of a ten-sided dice. The same is true for Option B, which yeilds either 3.85 tokens or 0.110 tokens depending on the roll of the dice.

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

	Option A	Option B	Your Choice A or B	
Decision 1	2.00 tokens if Dice is 1	3.85 tokens if Dice is 1	A	B
	1.60 tokens if Dice is 2-10	0.10 tokens if Dice is 2-10		

Please Choose Option A or Option B for the Decision Below.

40

	Option A	Option B	Your Choice A or B
Decision 1	2.00 tokens if Dice is 1	3.85 tokens if Dice is 1	A B
	1.60 tokens if Dice is 2-10	0.10 tokens if Dice is 2-10	

[*Page 7*] This concludes Game 3. Please wait for the experimenter to come.

4. Double Auction

[*Page 1*] **Instructions**

In this game there are two players: Seller and Buyer. You will be randomly matched with another person in this room. You will not be told the identity of the person that you are matched with and the other person will not know your identity. One of you will be randomly assigned the role of Seller and the other the role of Buyer. If this game is selected for payment, your payoffs will be determined by the decisions made by the Buyers and Sellers.

In this game, 1 token is equal to $1.

[*Page 2*] Every Seller can sell at most one unit of a fictitious good.The minimum assigned price at which the Seller can sell the unit of this good in any period will appear on the top left corner of the screen. The Buyer can buy at most one unit of the fictitious good. The maximum assigned price at which the Buyer can buy a unit of this good will appear on the top left corner of the screen.

Your assigned value will be private information. There will be 10 trading periods of 2 minutes each. Your assigned price will remain the same for all ten periods.

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

[*Page 3*] Before the game starts, the computer will randomly assign half of the participants the role of Buyers and the other half of the participants the role of Sellers. At the beginning of each trading period, two pieces of information will appear on the top left corner of your screen: your role (either Buyer or Seller) and your assigned price for the good. Please remember that your role and assigned price will remain the same for all ten trading periods. The first trading period will start when every player is done reading the instructions.

[*Page 4*] The goal of every trader is to maximize their payoff at every trading period. Thus, each Seller has to try to sell the good at the highest possible price, and each Buyer has to try to buy the good at the lowest possible price. Prices must be multiples of 0.5.

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

[*Page 5*] **Important remarks:**

Buyers cannot bid above their assigned maximum prices. That is, Buyers are only allowed to propose prices below or equal to their assigned prices for the good. If Buyers do not make a purchase, Buyers do not earn anything in the period. Similarly, Sellers cannot ask values below their assigned minimum prices. That is, Sellers are only allowed to propose prices above or equal to their assigned prices for the good. If Sellers do not make a sale, Sellers do not earn anything or incur any cost in that period.

[*Page 6*] A transaction is finalized when a Buyer accepts a Seller's offer, or when a Seller accepts a Buyer's bid. The Buyer and Seller making the deal are to drop out of the market, making no more bids, offers, or contracts for the remainder of that trading period. This process continues for up to two minutes depending

on the volume of trading.If Buyers or Sellers resubmit bids or offers, a transaction will automatically cancel all prior bids and offers made by the Buyers and Sellers involved. Bids and offers made by those who are not involved in the transaction do not have to be reentered. The highest bid and the lowest offer will be displayed at all times on your screen.

[*Page 7*] Your payoff will be equal to:

-the difference between the closing price and your assigned minimum price if you are a Seller;

-the difference between your assigned maximum price and the closing price if you are a Buyer.

Once you complete a transaction, your payoff for that period will appear on the screen.

[*Page 8*] Once all decisions have been made the game will end. We will calculate payoffs based on the decisions made. Again, you will be randomly paid for one of the games at the conclusion of the experiment.

[*Page 9. 10 Periods of auctions were conducted.*]

Period: 1

You are a: **SELLER**

Price of Good: [ ]

Trading will start soon!

[*Page 10*] This concludes Game 4.

5. Eckel and Grossman Gambling Task

[*Page 1*] **Instructions**

In this game, you will be selecting from one of six available gambles. The six different gambles will be listed on your GAMBLE SELECTION screen. You must choose one, and only one, of these gambles. To select a gamble, type the option that corresponds to that gamble into the form on the Decision Screen. Each gamble has two possible outcomes, High Amount or Low Amount, with equal probability of the event occurring. If this game is selected for payment, your compensation for the study will be determined by:

1) which of the six gambles you select; and

2) which of the two possible events occur.

[*Page 2*] Remember, every gamble has two possible outcomes that can occur with equal chance. At the end of this session, if this game is selected for payment, we will ask one of you to volunteer to draw a chip from a bag. The bag will contain two chips. One chip has an H on it representing High Amount. The other chip has an L on it representing Low Amount. If the volunteer pulls out the H chip, you earn the High Amount from the choice you picked. If the volunteer pulls out the L chip, then you earn the Low Amount from the choice you picked. Because there are only two chips, each chip has equal chance to be pulled out of the bag.
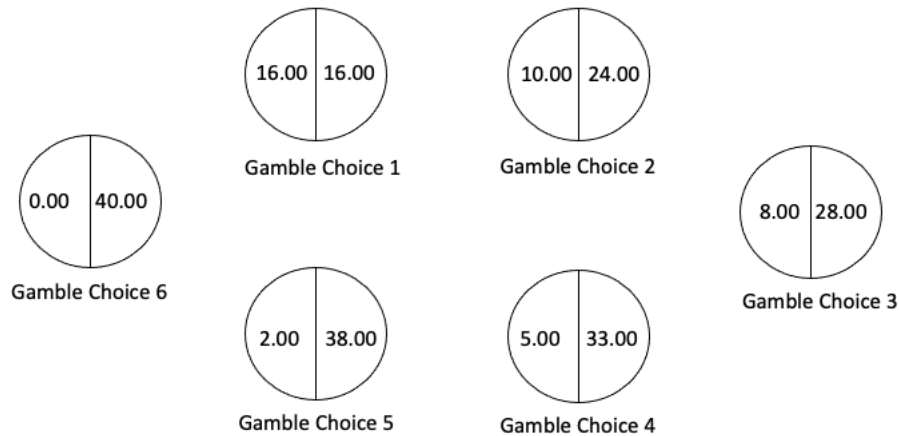
In this game, 2 tokens are equal to $1.

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

[*Page 3*] Once all decisions have been made the game will end. We will calculate payoffs based on the decisions made. Again, you will be randomly paid for one of the games at the conclusion of the experiment.

Directions: You must choose one and only one of these gambles. To choose a gamble, please type the gamble option that you prefer into the form.Select only one.

Gamble Choice 1 — 16.00 | 16.00

Gamble Choice 2 — 10.00 | 24.00

Gamble Choice 6 — 0.00 | 40.00

Gamble Choice 3 — 8.00 | 28.00

Gamble Choice 5 — 2.00 | 38.00

Gamble Choice 4 — 5.00 | 33.00

Decision

When you are ready please enter the gamble option (1,2,3,4,5, or 6) that you prefer. Remember, there are no right or wrong answers; you should just choose the option that you like best.

[Page 6] This concludes Game 1. Please wait for the experimenter to come.

6. Public Goods game

[Page 1] **Instructions**

In this game, you will participate in a total of 12 periods (2 practice and 10 real). In each period, you will be randomly assigned to a group of 4 members. Each member will be endowed with 100 tokens and must decide how to divide the tokens between two accounts:

1) Private Account

2) Public Account

In this game, 1 token is equal to 10 cents ($0.10).

[Page 2] The composition of your group will change every period. Each period, you will be randomly reassigned to a new group of 4 members. At no point in the experiment will the identities of the other group members be revealed to you, nor will your identity be revealed to them. In other words, the group members will remain anonymous to one another. You will be endowed with 100 tokens in every period and must decide how many tokens to invest in the private account and how many tokens to invest in the public account. Information about the two accounts is presented in the next four screens.

[Page 3] Private Account: Every token you invest in the private account will yield you a return of 10 cents. The other members in your group will not be affected by your investment in the private account. Here are a few examples to illustrate:

Example 1: Suppose you invest 100 tokens in the private account. Then you will get 1,000 cents (or $10.00) from this account and the other members of your group will not be affected for that period.

Example 2: Suppose you invest 50 tokens in the private account. Then you will get 500 cents (or $5.00) from this account and the other members of your group will not be affected for that period.

Example 3: Suppose you invest 0 tokens in the private account. Then you will get 0 return from this account and the other members of your group will not be affected for that period.

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

[*Page 4*] Public Account: Every token you invest in the public account will yield a return of half a cent to each member of your group. Also, every token that any of your group members invests in the public account will yield a return of half a cent to each member of your group.

This means that your return from the public account will depend on the total number of tokens that you and the other members of your group invest in this account. The more the group invests in the public account, the greater the return to each member of the group from this account.

Here are a few examples to illustrate:

Example 1: Suppose you invest 0 tokens in the public account and the other three members of your group invest a total of 200 tokens in the public account. Then, the total number of tokens invested by your group in the public account is 200 which means that every member of your group earns $200 \times 0.5 = 100$ tokens $\times$ 10 cents $= 1,000$ cents (or $10.00) from the public account for that period.

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

[*Page 5*] **Important remarks:**

Example 2: Suppose you invest 100 tokens in the public account and the other three members of your group invest a total of 0 tokens in the public account. Then the total number of tokens invested by your group in the public account is 100 which means that every member of your group earns $100 \times 0.5 = 50$ tokens $\times$ 10 cents $= 500$ cents (or $5.00) from the public account for that period.

Example 3: Suppose you invest 100 tokens in the public account and the other three members of your group invest a total of 300 tokens in the public account. Then the total number of tokens invested by your group in the public account is 400 which means that every member of your group earns $400 \times 0.5 = 200$ tokens $\times$ 10 cents $= 2,000$ (or $20.00) from the public account for that period.

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

[*Page 6*] Your decisions and earnings in every period are confidential. This means that you will not be given information about the investment decisions or earnings of any of your group members, nor will they be given information about your investment decisions or earnings. So you must make your decision without knowing what the other members in your group are deciding.

After each period, the only information you will be given is:

1) Number of tokens you invested in the private and public accounts

2) The total number of tokens invested by your group (including you) in the public account

3) Your earnings for that period

At the end of the session, if this game is selected for payment, 1 of the 10 real periods will be randomly selected as binding. We will ask one of you to volunteer to draw a chip from a bag. The bag will contain ten chips, one for each period in this game.

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

[*Page 6*] Once all decisions have been made the game will end. We will calculate payoffs based on the decisions made. Again, you will be randomly paid for one of the games at the conclusion of the experiment.

[*Page 7. 2 practice rounds were conducted.*]
Practice Round 1
Please enter your contribution to the private and public accounts and click NEXT. Remember, that your combined contribution to both the private account and public account must equal 100 tokens.
Group 1
Your endowment is 100 tokens
Your contribution (tokens) in the private account:
Your contribution (tokens) in the public account:

[*Page 8. 2 practice rounds were conducted.*]
You are in Group 1
Your Profit

Your contribution (tokens) in your private account:	
Your contribution (tokens) into the public account:	
Total group investment (tokens) into the public account	
Your earnings for this period	$

[*Page 9. 10 periods of public goods game are conducted.*]
Period 1
Please enter your contribution to the private and public accounts and click NEXT. Remember, that your combined contribution to both the private account and public account must equal 100 tokens.
Group 1
Your endowment is 100 tokens
Your contribution (tokens) in the private account:
Your contribution (tokens) in the public account:

[*Page 10. 10 periods of public goods game are conducted.*]
Period 1
You are in Group 1
Your Profit

Your contribution (tokens) in your private account:	
Your contribution (tokens) into the public account:	
Total group investment (tokens) into the public account	
Your earnings for this period	$

[*Page 12*] This concludes Game 2. Please wait for the experimenter to come.

7. Ultimatum game

[*Page 1*] **Instructions**

In this game there are two players: Player 1 and Player 2. You will be randomly matched with another person in this room. You will not be told the identity of the person that you are matched with and the other person will not know your identity. One of you will be randomly assigned the role of Player 1 and the other the role of Player 2. Player 1 begins with 10 tokens and Player 2 begins with 0 tokens. There will be one period only.

<div align="center">In this game, 1 token is equal to $1.</div>

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

[*Page 2*]

Player 1's Decision:

Player 1 will be asked to propose to split 10 tokens between him/herself and Player 2. Player 1 can choose any integer amount between 0 and 10 tokens.

Player 2's Decision:

Player 2 will be told the offer made by Player 1 and asked to choose to either **accept** or **reject** the offer.

If Player 2 **accepts** the offer, both Player 1 and Player 2 receive the amounts specified in the **accepted** offer. If Player 2 **rejects** the offer, both players receive zero earnings for this game.

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

[*Page 3*] Once all decisions have been made the game will end. We will calculate payoffs based on the decisions made. Again, you will be randomly paid for one of the games at the conclusion of the experiment.

[*Page 4*] You are Player 1./You are Player 2.

[*Page 5. Player 1's decision.*]

<div align="center">You are assigned as Player 1.</div>

<div align="center">Now, please select the number of tokens you propose to send to Player 2.</div>

Number of tokens to transfer to Player 2										
0	1	2	3	4	5	6	7	8	9	10

[*Page 6. Player 2's decision.*]

Player 1 was endowed **10 tokens**.

Player 1 decided to offer you [ ] and keep [ ] tokens.

You can now decide, if you want to **accept** or **reject** the offer. If you **reject** the offer, both of you will receive **0 tokens** in this game.

Now, please select if you **accept** or **reject** Player 1's offer.

<div align="center">

Accept

</div>

<div align="center">

Reject

</div>

[*Page 7*] This concludes Game 3. Please wait for the experimenter to come.

8. Cheating game

[*Page 1*] **Instructions**

In this game, you will see 5 numbers on the screen. You will determine your own payoff by recording the first number that you see. Then, you will move to the next period. You will play this game for 10 periods.

In this game, 1 token is equal to $1.

In other words, if you report seeing "0" first, you receive 0 tokens. If you report seeing "2" first, you receive 2 tokens, or if you report seeing "4" first, you receive 4 tokens.

[*Page 2*] Each screen will present to you 5 numbers. The numbers are 0, 2, 4, 6, and 8 in different orders. Your task is to report the first number that you see and record the number on the next screen. The position of the numbers on the screen are determined by a computer generated output. Before this session began, we called a participant to monitor the randomization process.

Please raise your hand if you have any questions, otherwise please click NEXT to continue.

[*Page 3*] At the end of this session, if this game is selected for payment, we will ask one of you to volunteer to draw a chip from a bag. The bag will contain ten chips, one for each of the periods in this game. Each period has an equal chance of being selected.

Once all decisions have been made the game will end. We will calculate payoffs based on the decisions made. Again, you will be randomly paid for one of the games at the conclusion of the experiment.

[*Page 4. 10 periods of cheating games were conducted.*]

Period 1

Please remember the first number.

6   8   0   4   2

[*Page 5. 10 periods of cheating games were conducted.*]

Period 1

Please record the first number that you saw.

☐

When you are finished, please click "NEXT" to continue.

[*Page 6*] This concludes Game 4.