Eager beavers v. lazy slugs: Selection effects in experiments with social preferences

Catherine Eckel[1], Rick K. Wilson[2*], Sora Youn[1]

[1*]Department of Economics, Texas A&M University, College Station, 77845, Texas, USA.
[2]Department of Political Science, Rice University, Houston, 77005, Texas, USA.

*Corresponding author(s). E-mail(s): rkw@rice.edu;
Contributing authors: ceckel@tamu.edu; sorayoun@tamu.edu;

Abstract

We ask whether social preferences measured in subjects who come to the laboratory when invited are systematically different from those of subjects who only respond when an online option is available. Subjects participated in two types of third-party (other-other) dictator games and a trust game, either in the lab or on-line. In the third party dictator games, the dictator divides $20 between two other individuals, one of whom is a member of their in-group. (We also varied types of in-group between a real group and an artificial group.) In the trust game, the first-mover decides how much of the endowment to send to the second-mover. The second-mover receives the amount sent tripled by the experimenter and decides how much to send back to the trustee. Across all the games, we find no statistically significant differences in social preferences measured in-lab and on-line.

Keywords: online experiment, methodology, social preferences

JEL Classification: C81 , C90

Statements and Declarations

Acknowledgments

This study was funded by the National Science Foundation (for Wilson: SES-1534403 and SES-2027556; for Eckel: SES-1534411 and SES- 2027548), and has received IRB approval at Texas A&M and Rice University. Protocols, data and Supporting Information is available at: **. Pre-registered hypotheses are available at: https://doi.org/10.17605/OSF.IO/PZ5WG.

Competing Interests

The authors declare they have no competing interests.

1 Introduction

When COVID-19 disrupted laboratory experiments around the world, many researchers turned to online experiments. While maintaining control over human subjects in the laboratory is considered the gold standard for conducting experiments, circumstances often lead to relaxing that control. Now that the pandemic is less of a threat scholars have the luxury of staying online or returning to the lab. Subjects, too, have the choice to participate in laboratory experiments or opt for experiments carried out online. Does it matter if subjects self-select into their venue?

When measuring social preferences, this question is especially important. The lab may attract more socially motivated subjects, leading to a finding of more generous social preferences in the lab. On the other hand, going to the lab is a significant time commitment, which might tend to push results the other way. Those who participate online may translate their time savings into more generous behavior. Other factors, such as the closer observation of lab subjects, or uncertainty about the existence of a partner in an online setting, can also play a role. As researchers, how much should we worry about the impact on our results of self-selected subjects in lab and online settings?

In this study we invited all subjects who were part of a longitudinal panel to participate in a study measuring social preferences. Those who signed up first (the eager beavers) completed the task in a traditional laboratory setting. After the laboratory experiments were completed we recontacted the remaining members of the panel (the lazy slugs) and asked them to complete an online version using the same protocol. We fully expected that the online subjects would behave differently due in part to selection. To our surprise, we find no significant differences in behavior between the two groups of subjects. This should provide some reassurance to researchers when choosing one setting or another.

2 Motivation

It is well-known that online experiments are vulnerable to numerous threats to validity, including subject distraction, absence of experimenter monitoring, expectations about payments, beliefs about their counterparts, selective dropouts, and outside

consultation (Dandurand et al. (2008), Clifford and Jerit (2014), Eckel and Wilson (2006), Horton et al. (2011)). In addition, the online environment may increase perceived social distance, systematically dampening social preferences (Akerlof (1997)). Nevertheless, most studies conclude that online experiments are comparable to lab experiments (Horton et al. (2011), Brañas-Garza et al. (2018), Brañas-Garza et al. (2023)). We stress-test such results by asking whether this equivalence holds when subjects self-select into lab or online experiments.

When studying social preferences, evidence from prior studies generally support equivalence, but is somewhat mixed. Buso et al. (2021) conduct standard dictator, ultimatum, and public goods games and find no systematic differences in pro-social behavior across different settings: in-lab, online with video monitoring, and online without video monitoring. However, while Hergueux and Jacquemet (2015) find "strong parallelism" between online and lab behavior, they note that online participants display slightly more pro-social behavior than their lab counterparts. Prissé and Jorrat (2022) find that most behaviors are consistent between the lab and online venues. However, in a dictator game with a charity recipient, online participants are slightly more likely to give zero. They ascribe such a difference to social distance and experimenter monitoring. Generally, these studies indicate that there are only minor differences due to venue.

In these experiments, researchers are careful to use individuals taken from the same subject pool, who participate at roughly the same time and are randomly assigned to the type of venue. Subjects are unable to choose how they would like to participate. If they are able to choose when and where they participate, will that affect the findings?

3 Experimental Design and Procedure

A random sample of two-thirds of Rice University's 2016 entering freshman class was recruited prior to arriving on campus and participated in an on-line experiment as part of a longitudinal panel study. A total of 553 of the 661 contacted completed Phase 1 of the study (992 matriculated). Three months after starting classes, those who completed Phase 1 were invited to participate in Phase 2 of the study. A total of 521 subjects participated beginning in early November 2016.

In Phase 2 subjects were recruited in two ways. First, all panelists were sent recruitment emails asking them to sign up for an in-lab experiment. From November 2, 2016 through November 23, 2016 a total of 22 lab sessions were run with 236 subjects. Second, the remaining panelists who had not yet participated were recruited for an online experiment, which was open from November 28, 2016 – April 28, 2017.[1] This yielded another 285 subjects. The experimental interface was identical for both the lab and online experiments (see the Supporting Information (SI) - Section 2). Note that subjects were not randomly assigned to one form of participation or the other.

In this paper, we focus on two third-party dictator games and a standard trust game (details are in the SI - Section **??**). In the third-party dictator games, the dictator

[1]The online portion of the experiment coincided with the end of the academic term. Over 93 percent of the subjects completed the online study by February 1, 2017.

divides $20 between two other individuals, one of whom is a member of their in-group, and the other of whom is not. Dictators were paid a fixed fee of $5.00 for the decision. The third-party dictator games differed from one another in that the ingroup recipients were either from their own residential college relative to someone who is in a different residential college (a *real* group) or part of a *minimal* ingroup relative to someone who is in the outgroup. Subjects completed the two games in random order, and we control for the order effects in our discussion (for further details about these treatments, see Eckel et al. (2022)).

The trust game is widely accepted as a measure of interpersonal trust and trustworthiness (Berg et al. (1995)). Each actor earned an initial endowment of $10 for completing a 40-item risk survey. In the game, the first-mover (the truster) decides how much of the endowment to send to the second-mover. The second-mover (the trustee) receives an amount equal to three times the amount sent (tripled by the experimenter) and decides how much to send back to the trustee. Subjects used the strategy method to decide how much to return conditional on all possible amounts sent.

All subjects completed all of the tasks, and were informed that their counterparts were also participants in the study. One of the tasks was randomly selected for payment. For the dictator games subjects were randomized to role (the dictator, the in-group member, or the out-group member). In-group and out-group members were paid based on the allocation of the $20. For the trust game, subjects played both roles - truster and trustee - and at the end of the experiment were randomly assigned to one position. Thus both games have a "role uncertainty" design (Iriberri and Rey-Biel (2011)). Subjects were not told which task was paid until the end of the experiment and the tasks were randomly chosen for each subject. All of the randomization and matching to positions was computerized. For subjects in the lab, matching was within session. For subjects who were online, matching was with others participating online and payments were delayed until the end of the experiment. Subjects spent less than 30 minutes, either in the lab or on-line and earned an average of $21.37.

4 Experimental Results

The analysis proceeds as follows. Using the third-party dictator games, we compare in-group favoritism in the lab and online for the two games (real and minimal groups). Next, we turn to the trust game data and focus on two measures. The first is the amount sent by the first mover (a measure of trust). The second is the average percentage returned (reciprocity). Under the strategy method subjects specified how much they would return contingent on each whole dollar that could be sent. The percentage returned is calculated for each strategic choice and the average per subject is used as the measure.

First, we find that there are few differences between the type of subjects opting for the laboratory and those taking up the online option. Table 1 uses several standard demographic measures and we find that the eager beavers are very similar to the lazy slugs. There is balance between males and females. Asians are more likely to show up in the lab than Caucasians. When we look at a measure of risk aversion collected prior to matriculation, we find no difference between the two sets of participants.

The same is true for a measure of time preferences.[2] There are no differences in GPA measured in several ways. Finally, across the five personality inventory items, we find no significant differences. The differences we note disappear when adjusting for multiple hypothesis testing Westfall and Young (1993). All-in-all the two groups are balanced across multiple measures.

Table 2 presents the social preferences measured in this study and shows the mean differences and p-values of t-tests. The first two rows show giving to an ingroup member (relative to a non-ingroup-member) from the $20 budget. Both rows show ingroup favoritism (amounts greater than $10, more than half of the budget, are sent to the ingroup members). Adjusting for multiple hypothesis testing leads to no significant differences being detected.[3] The last two rows point out there are no significant differences in trust or reciprocity.[4]

Figure 1 graphs the mean for each incentivized measure and includes the 95 percent confidence interval. This figure illustrates what is detailed in Table 2. There is ingroup bias in the dictator games and the effect is true for both laboratory and online subjects. There are no differences for the trust game.

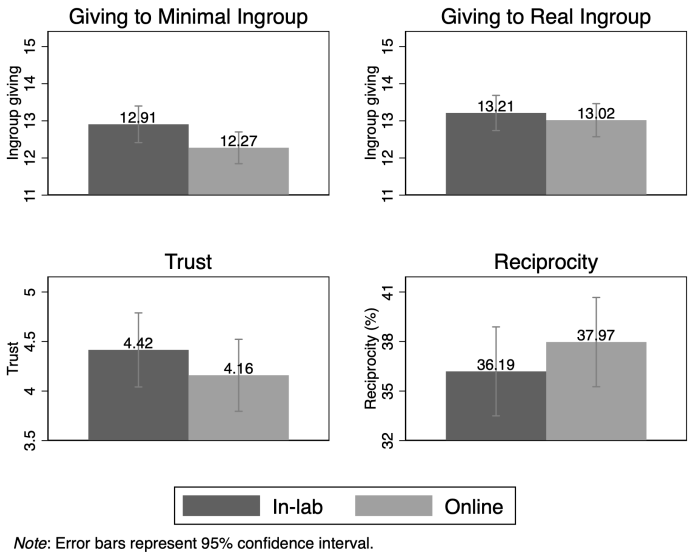


Fig. 1 Means for Social Preference Measures

[2]These incentivized measures were collected prior to matriculation. A discussion of these measures is reported in Eckel et al. (2023).

[3]We are also concerned with ordering effects for the dictator games. Analysis reported in the SI - section ?? shows there is no difference when adjusting for multiple hypothesis testing.

[4]In the online study, two observations were lost due to a programming error. Hence the differences in n's between the dictator games and the trust game.

Table 1 Sample Characteristics

	In-lab		Online		Difference		H_0: In-lab = Online	
	Mean	SE	Mean	SE	Mean	SE	Unadj.p-val	Adj. p-val
Demographics								
Female	0.53	0.03	0.48	0.03	0.04	0.04	0.31	0.98
Ethnicity:								
Black	0.06	0.01	0.05	0.01	0.00	0.02	0.90	1.00
Asian	0.33	0.03	0.22	0.02	0.11	0.04	0.01	0.08
White	0.28	0.03	0.39	0.03	-0.11	0.04	0.01	0.16
Hispanic	0.14	0.02	0.16	0.02	-0.02	0.03	0.59	1.00
Citizen	0.83	0.02	0.87	0.02	-0.04	0.03	0.20	0.94
Political Inclination	2.46	0.06	2.44	0.05	0.02	0.08	0.77	1.00
Pre-matriculation measures								
Risk Aversion	3.21	0.11	3.18	0.09	0.03	0.14	0.86	1.00
Time Preference	2.49	0.08	2.59	0.08	-0.10	0.11	0.38	0.98
Academic achievement								
GPA in the first semester	3.61	0.03	3.55	0.03	0.06	0.04	0.18	0.93
Cumulative GPA by Fall 2021	3.62	0.02	3.59	0.02	0.03	0.04	0.37	0.98
Number of completed hours by Fall 2021	139.16	1.49	138.28	1.32	0.88	1.99	0.66	1.00
Short Form Personality Inventory								
Extraversion	4.12	0.11	3.94	0.10	0.18	0.15	0.23	0.95
Agreeableness	4.96	0.08	4.84	0.08	0.11	0.11	0.32	0.98
Conscientiousness	5.38	0.08	5.38	0.07	0.01	0.10	0.94	1.00
Emotional Stability	4.60	0.09	4.87	0.08	-0.28	0.12	0.03	0.30
Openness to New Experiences	5.34	0.07	5.26	0.06	0.08	0.09	0.36	0.98

Note: Westfall and Young (1993) adjusted p-values are estimated using 10,000 bootstraps. Stata command *wyoung* (Jones et al. (2019)) is used. *Political Inclination* ranges from 1 (liberal) to 5 (conservative). Except for one who started in Spring 2017, our sample started the first (regular) semester in Fall 2016. 5 started in Summer Semester 2016.

Table 2 Means for Social Preference Measures

	In-lab	Online	Difference	Unadj.p-val	Adj. p-val
Giving to Minimal Ingroup	12.907	12.274	0.633	0.056	0.179
	(0.251)	(0.218)	(0.331)		
	n=236	n=285			
Giving to Real Ingroup	13.212	13.018	0.194	0.559	0.674
	(0.241)	(0.227)	(0.332)		
	n=236	n=285			
Trust	4.415	4.159	0.256	0.337	0.674
	(0.190)	(0.185)	(0.267)		
	n=236	n=283			
Reciprocity	36.193	37.965	-1.772	0.366	0.674
	(1.371)	(1.376)	(1.959)		
	n=236	n=283			

Note: Standard errors in parentheses. Adjusted p-values are calculated using Stata command *wyoung* (Jones et al. (2019)).

5 Conclusion

We conclude that online and in-lab measures of social preferences are robustly consistent for early and late-takers in an experiment where subjects in the online version are recruited from those who fail to sign up for the lab version. We find minor differences in demographics (Asians are more likely, and Whites less likely, to enroll in a lab setting), and no significant differences in behavior in the two settings. This result should be reassuring for those who are concerned that online measures of social preferences are fundamentally different from their in-lab versions. Both yield similar patterns and outcomes. Once they engage with the experiment, eager beavers and lazy slugs are equally trusting and trustworthy, and favor their ingroup members to the same extent.

References

Akerlof, G. 1997. Social distance and social decisions. *Econometrica 65*(5): 1005–1028 .

Berg, J., J. Dickhaut, and K. McCabe. 1995. Trust, reciprocity, and social history. *Games and economic behavior 10*(1): 122–142 .

Brañas-Garza, P., V. Capraro, and E. Rascon-Ramirez. 2018. Gender differences in altruism on mechanical turk: Expectations and actual behaviour. *Economics Letters* 170: 19–23 .

Brañas-Garza, P., D. Jorrat, A.M. Espín, and A. Sánchez. 2023. Paid and hypothetical time preferences are the same: Lab, field and online evidence. *Experimental Economics 26*(2): 412–434 .

Buso, I.M., D.D. Cagno, L. Ferrari, V. Larocca, L. Lorè, F. Marazzi, L. Panaccione, and L. Spadoni. 2021, December. Lab-like findings from online experiments. *Journal of the Economic Science Association 7*(2): 184–193. https://doi.org/10.1007/s40881-021-00114- .

Clifford, S. and J. Jerit. 2014. Is there a cost to convenience? an experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political Science 1*(2): 120–131 .

Dandurand, F., T. Shultz, and K. Onishi. 2008, 05. Comparing online and lab methods in a problem-solving experiment. *Behavior research methods* 40: 428–34. https://doi.org/10.3758/BRM.40.2.428 .

Eckel, C., N. Sinha, and R. Wilson. 2023. The evolution of preferences and charitable giving: a panel study of the university years. *Oxford Economic Papers* 0: 1–21. https://doi.org/10.1093/oep/gpad030 .

Eckel, C. and R. Wilson. 2006. Internet cautions: Experimental games with internet partners. *Experimental Economics* 9: 53–66. https://doi.org/10.1007/s10683-006-4307-4 .

Eckel, C., R. Wilson, and S. Youn. 2022. In-group favoritism in natural and minimal groups. *Economics Letters* 219: 110794 .

Hergueux, J. and N. Jacquemet. 2015, June. Social preferences in the online laboratory: a randomized experiment. *Experimental Economics 18*(2): 251–283. https://doi.org/10.1007/s10683-014-9400-5 .

Horton, J., D. Rand, and R. Zeckhauser. 2011, September. The online laboratory: conducting experiments in a real labor market. *Experimental Economics 14*(3): 399–425. https://doi.org/10.1007/s10683-011-9273-9 .

Iriberri, N. and P. Rey-Biel. 2011. The role of role uncertainty in modified dictator games. *Experimental Economics* 14: 160–180 .

Jones, D., D. Molitor, and J. Reif. 2019, 08. What do Workplace Wellness Programs do? Evidence from the Illinois Workplace Wellness Study. *The Quarterly Journal of Economics 134*(4): 1747–1791. https://doi.org/10.1093/qje/qjz023. https://academic.oup.com/qje/article-pdf/134/4/1747/32666232/qjz023.pdf .

Prissé, B. and D. Jorrat. 2022. Lab vs online experiments: No differences. *Journal of Behavioral and Experimental Economics* 100: 101910. https://doi.org/https://doi.org/10.1016/j.socec.2022.101910 .

Westfall, P.H. and S.S. Young. 1993. *Resampling-Based Multiple Testing: Examples and Methods for P-value Adjustment.*

Supporting Information: Eager beavers v. lazy slugs: Selection effects in experiments with social preferences.

1 Sample

The sample for this project is based on individuals who were part of a longitudinal panel study. The initial sample was drawn from a random sample of two-thirds of Rice University's entering freshman class (992 students). These subjects were contacted prior to arriving on campus in the late summer of 2016 (from July 14 – August 12, 2016). A total of 553 of the 661 contacted students completed this portion of the study. The study was carried out online.

The current project recontacted the 553 students who completed the initial study. All subjects were invited to participate in an in-lab experiment and asked to sign up for a particular time slot (see the text of the email in Section 4).

Recruitment for the laboratory experiments took place at the end of the academic semester from November 2, 2016 through November 23, 2016. This resulted in a total of 22 lab sessions with 236 subjects. Those who did not participate in the lab were then solicited to participate online from from November 28, 2016 – April 28, 2017. This yielded another 285 subjects. Of those subjects participating online, over 82 percent completed the study prior to classes beginning on January 8, 2017. Only 9 subjects completed the study in March or April 2017. It is important to note that subjects were not randomly assigned to one form of participation or the other. The eager beavers were those who signed up for the in-lab experiments. The lazy slugs were those who waited until later.

2 Task Descriptions

In the current study subjects faced 5 distinct tasks: a 40-item risk instrument for which subjects were paid $10 to complete; a third-party dictator game (used in this study); a trust game; a second third-party dictator game (used in this study); and a 19-item demographic survey. Tasks 1, 3 and 5 were presented in fixed order. The third-party dictator games had either a strong group (the subject's own college) or a minimal group and were randomly ordered at the level of the subject. As noted, subjects were given a fixed amount ($10) for filling out a questionnaire. This was designed to serve

as money that was used in the trust game. One of the 3 incentivized tasks was randomly chosen for payment.

Subjects read the following instructions before beginning making a choice in the (strong) third-party dictator game. In this game subjects decided how to allocate a $20 budget between someone from their college or someone from a different college. All students matriculating at Rice are randomly assigned to a residential college. They live in those colleges for their four years at Rice. College attachments are very strong.

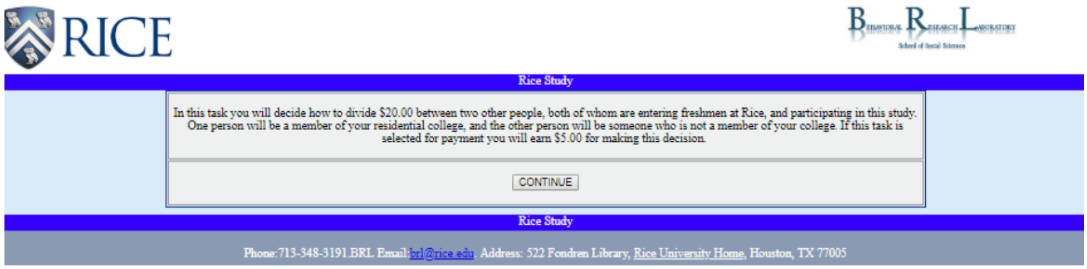


Figure 1: Instructions for (Strong) Third-Party Dictator Game

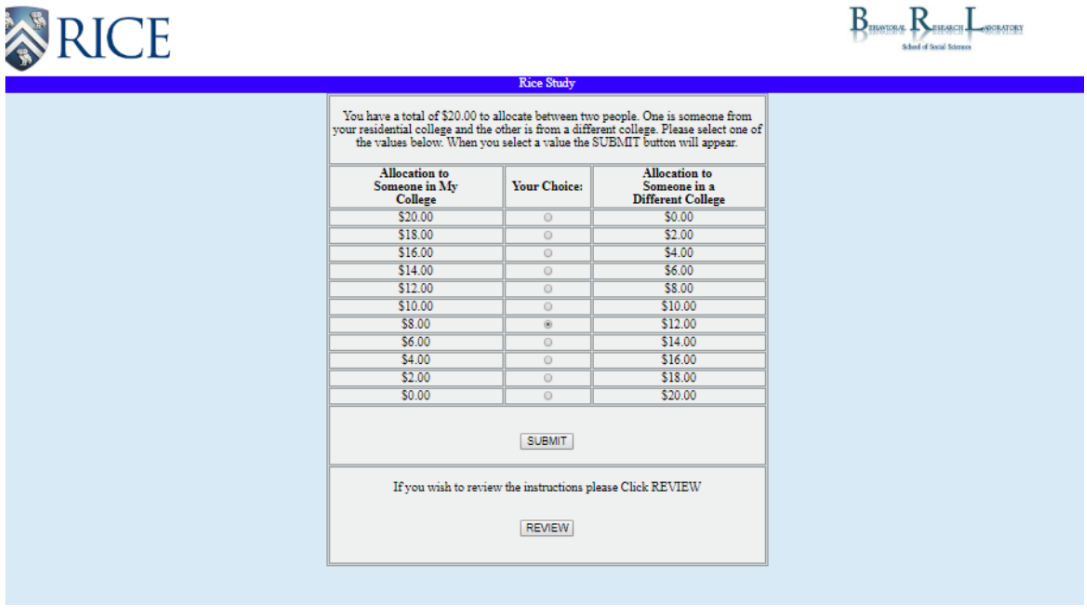


Figure 2: Decision Screen for (Strong) Third-Party Dictator Game

The (minimal) third-party dictator game was similar. Before making a choice subjects were shown a screen with dots for four seconds. asked to estimate how many dots appeared on their screen were told how they would be assigned to groups. Those instructions are reproduced here. Figure 10 notes the decision screen viewed by subjects in the minimal group condition.

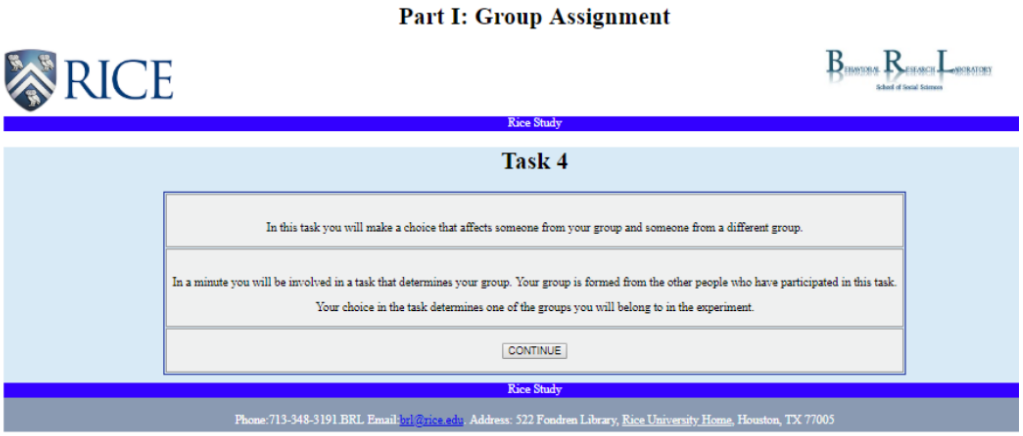


Figure 3: Phase 2 page 1 instructions for minimal group assignment

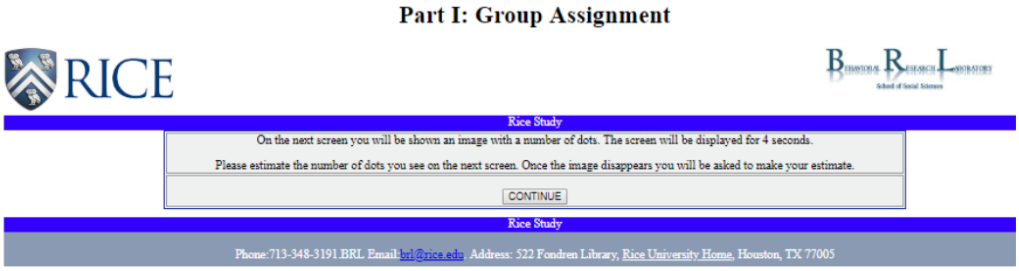


Figure 4: Phase 2 page 2 instructions for minimal group assignment

Figure 5: Phase 2 page 3 instructions for minimal group assignment. This is the dot estimation task that they faced.

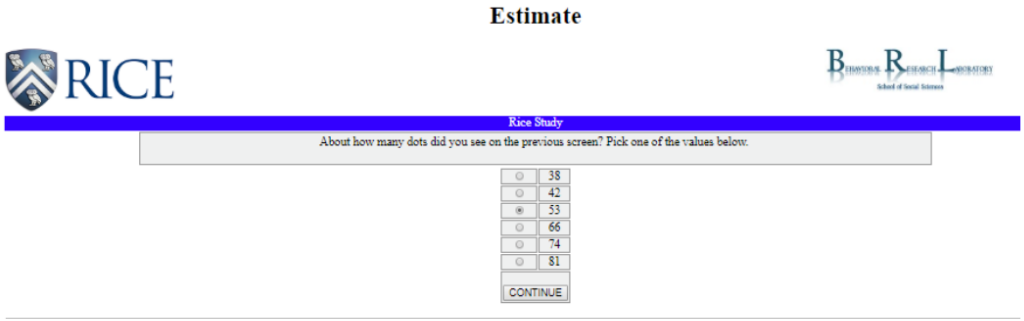


Figure 6: Phase 2 page 4 instructions for minimal group assignment. On this screen subjects guessed the number of dots from the previous page.

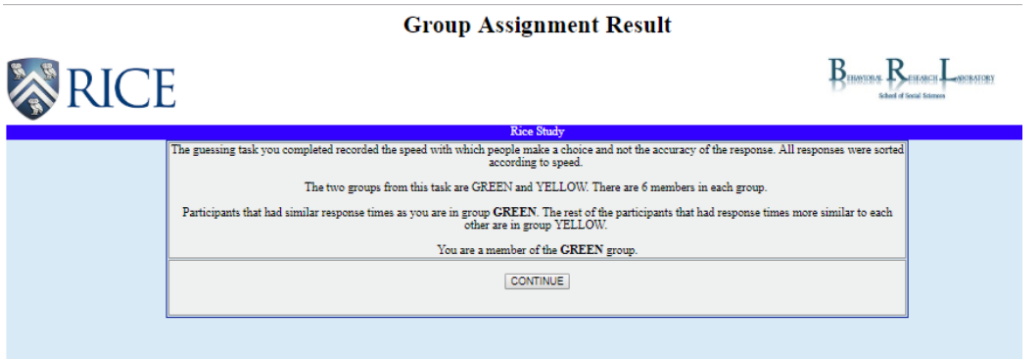


Figure 7: Phase 2 page 5 instructions for minimal group assignment. On this screen subjects were assigned to their group.

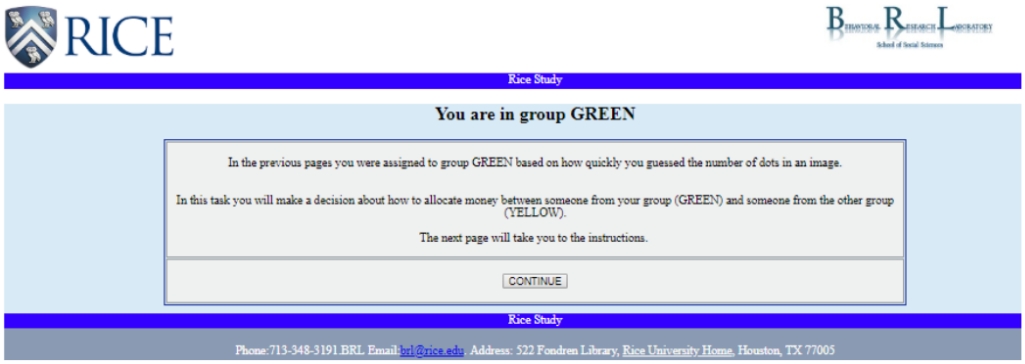


Figure 8: Phase 2 page 6 instructions for minimal group assignment

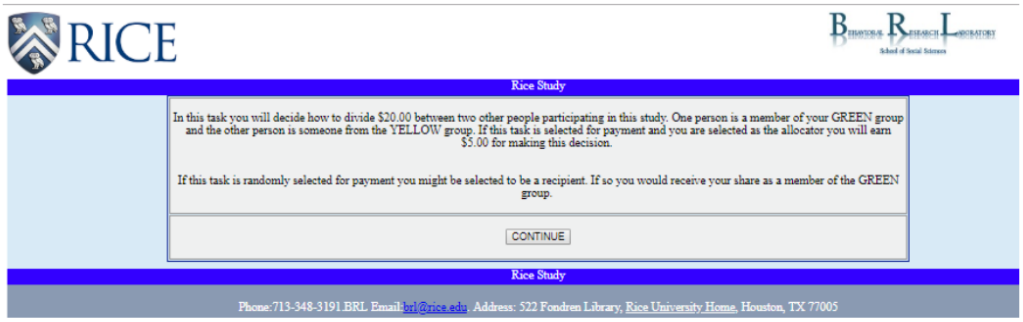


Figure 9: Phase 2 page 7 instructions for minimal group assignment

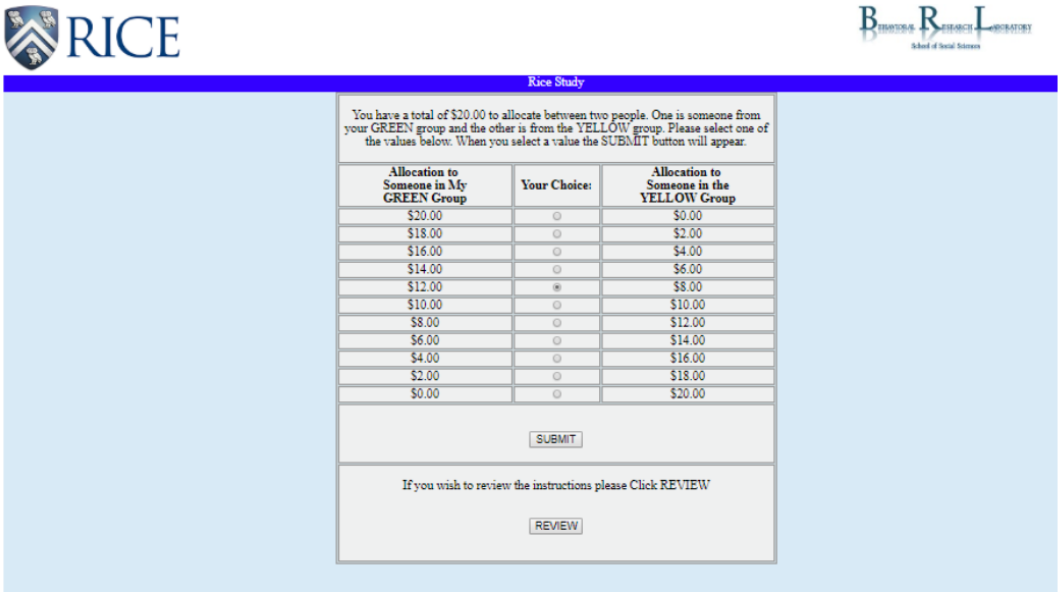


Figure 10: Decision Screen for Phase 2 Minimal Group Third-Party Dictator Game

Screen shots for the Trust Game.

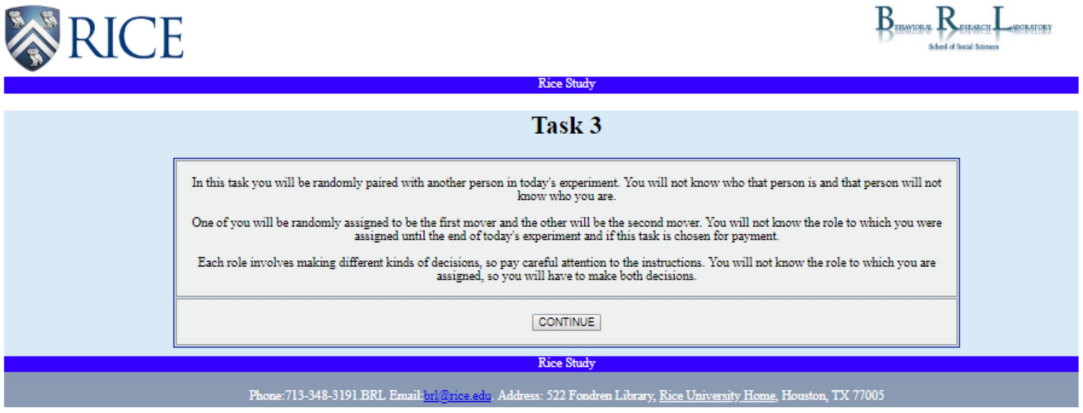


Figure 11: Introduction to the Trust Game

Figure 12: Trust game instructions page 2

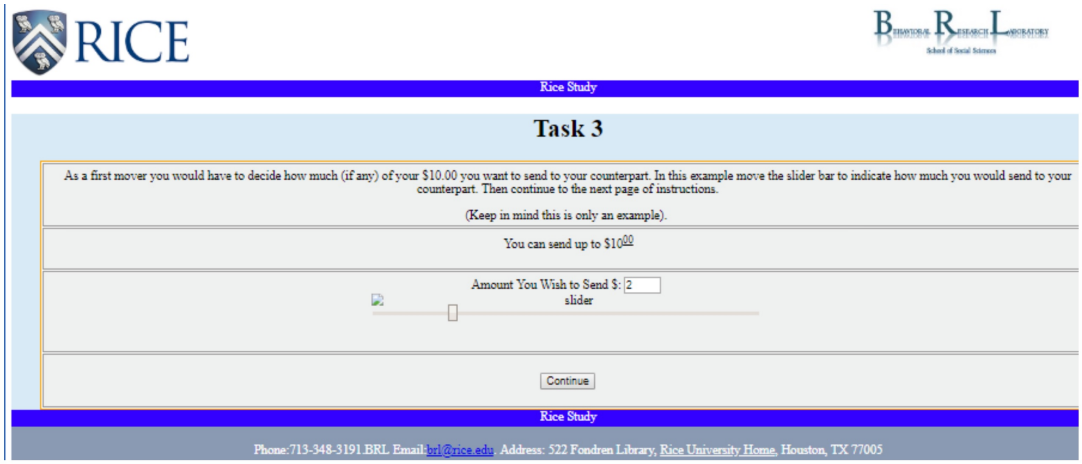


Figure 13: Trust game instructions page 3

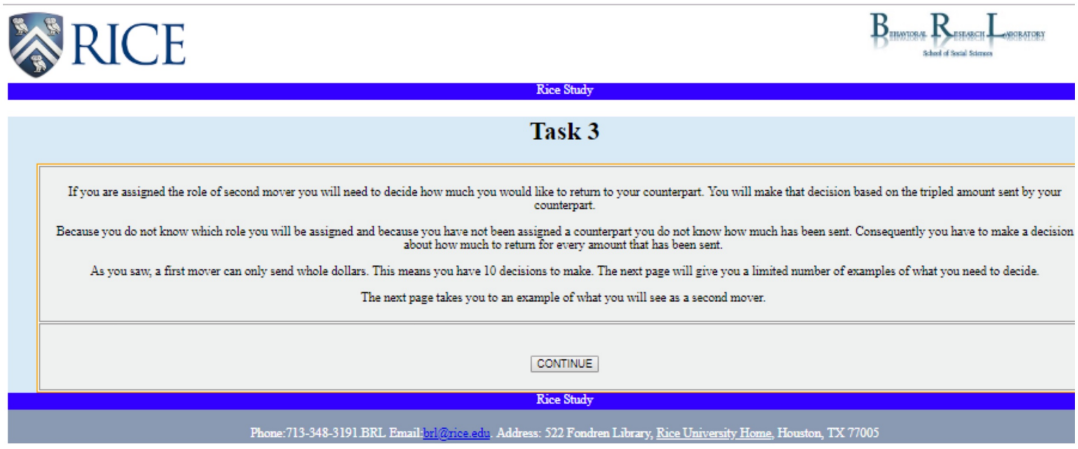


Figure 14: Trust game instructions page 4

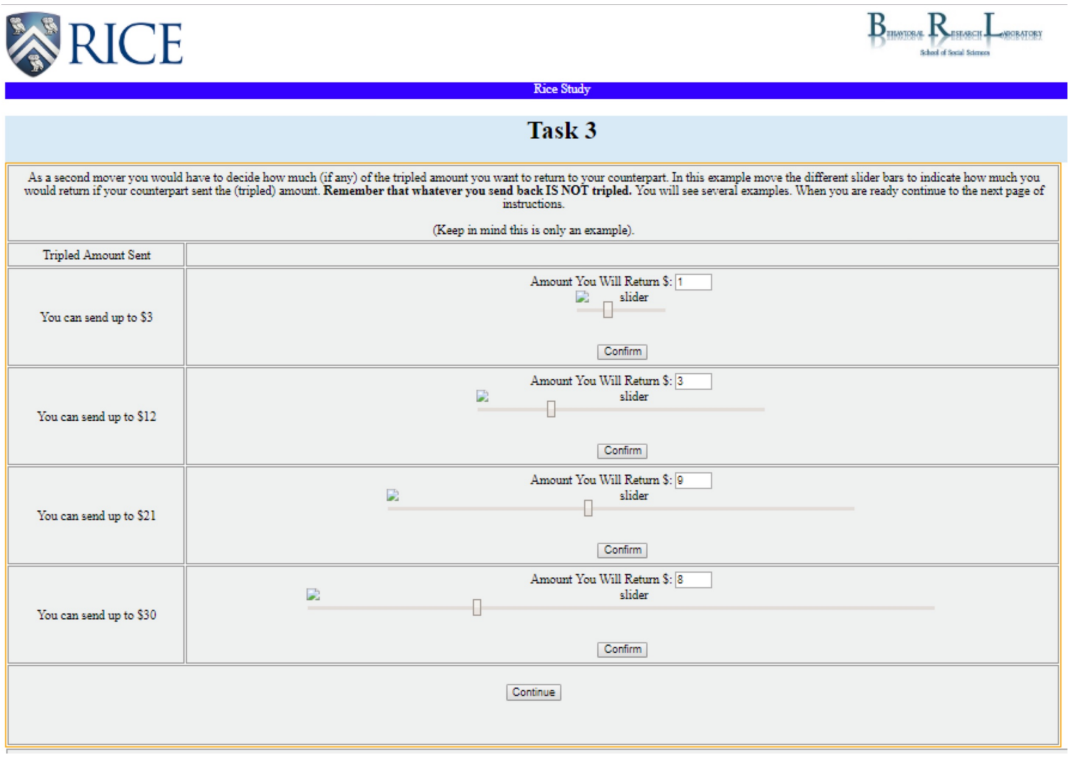


Figure 15: Trust game instructions page 5

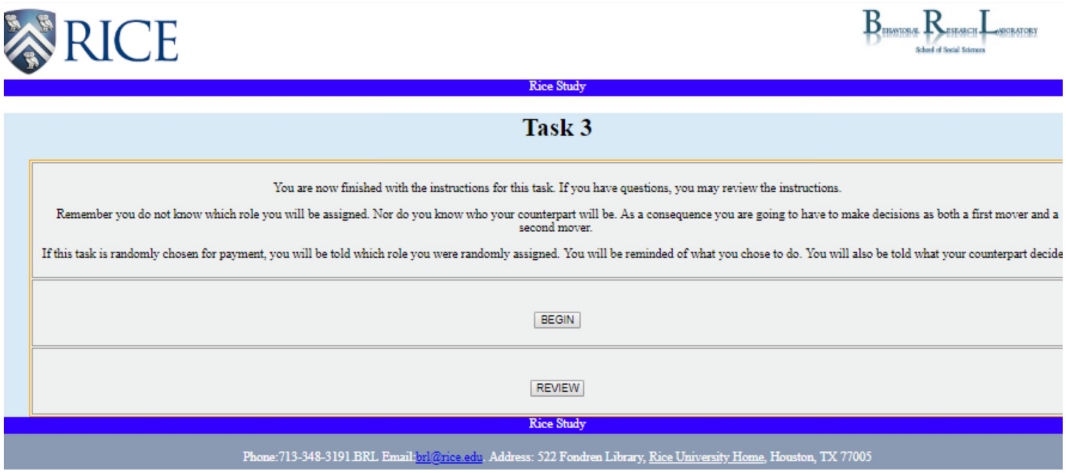


Figure 16: Trust game instructions page 6

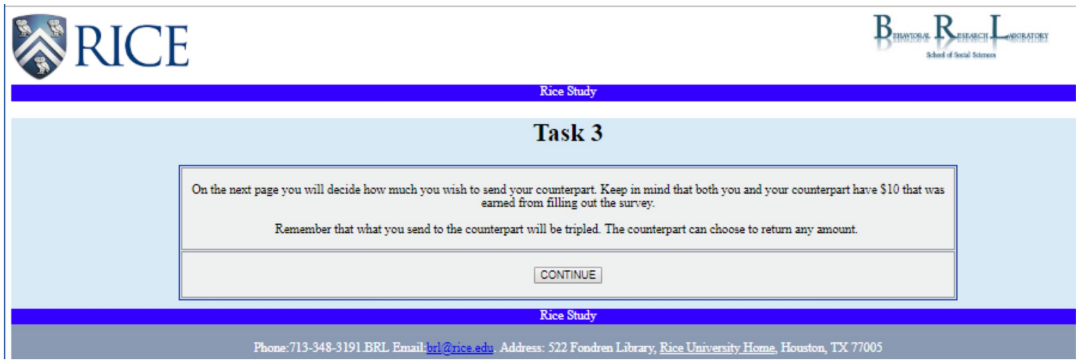


Figure 17: Trust game instructions page 7

Figure 18: Trust game decision

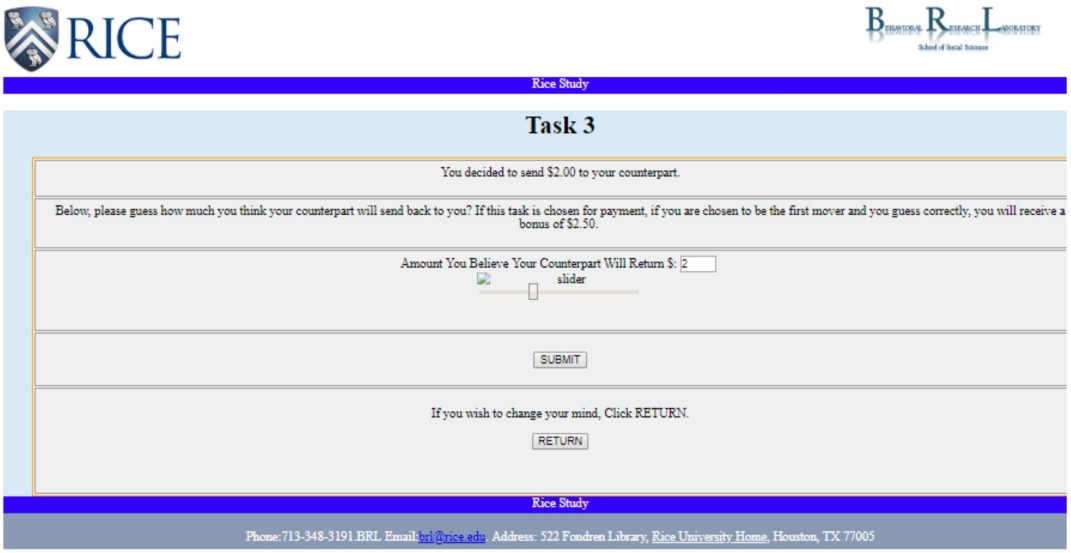


Figure 19: Trust game beliefs elicitation

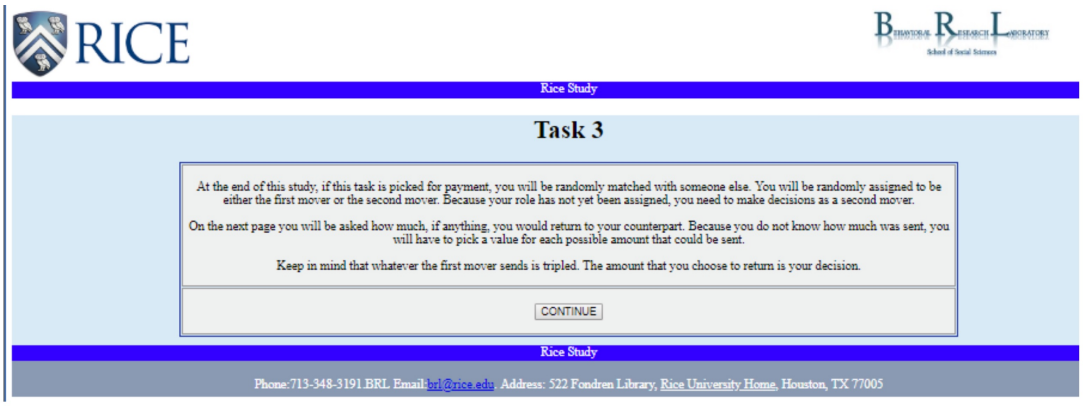


Figure 20: Trust game reciprocity

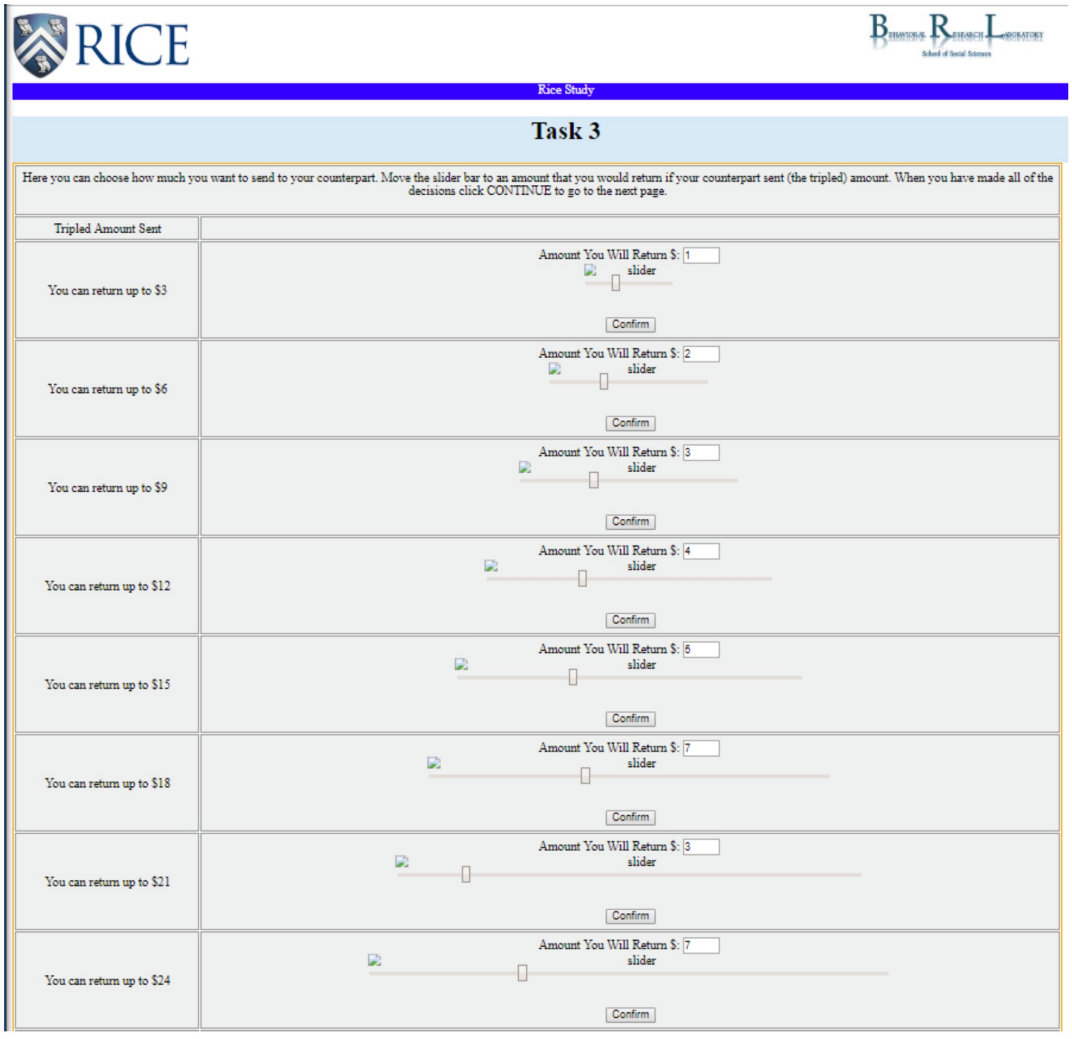


Figure 21: Trust game reciprocity decision

Figure 22: Trust game reciprocity decision

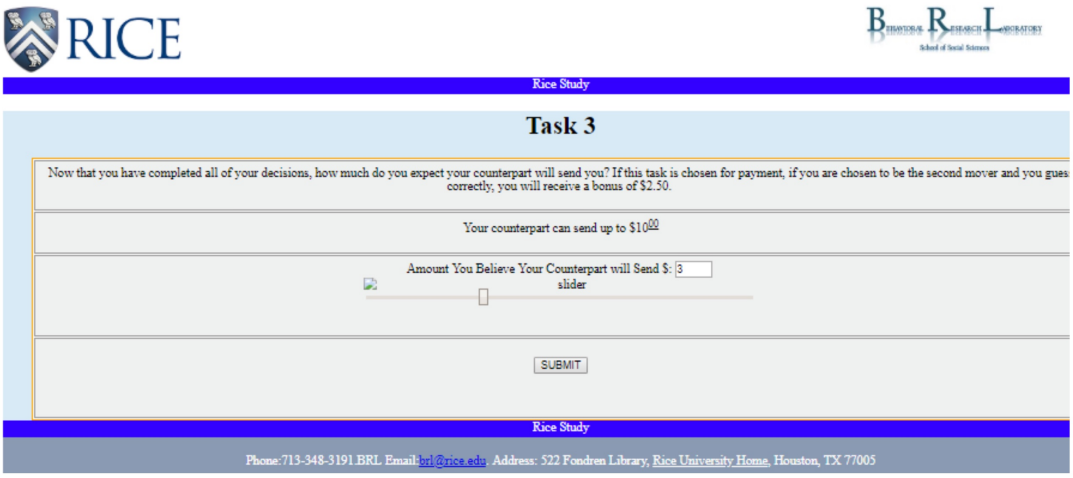


Figure 23: Trust game reciprocity beliefs

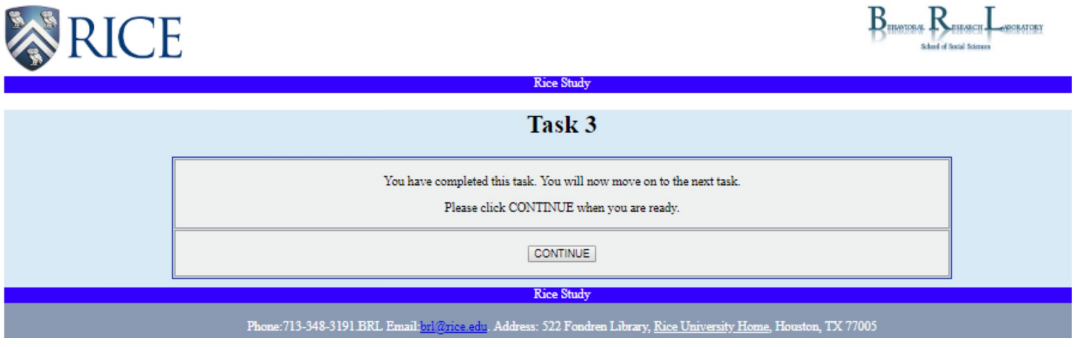


Figure 24: End of the Trust Game

3 Additional Analysis

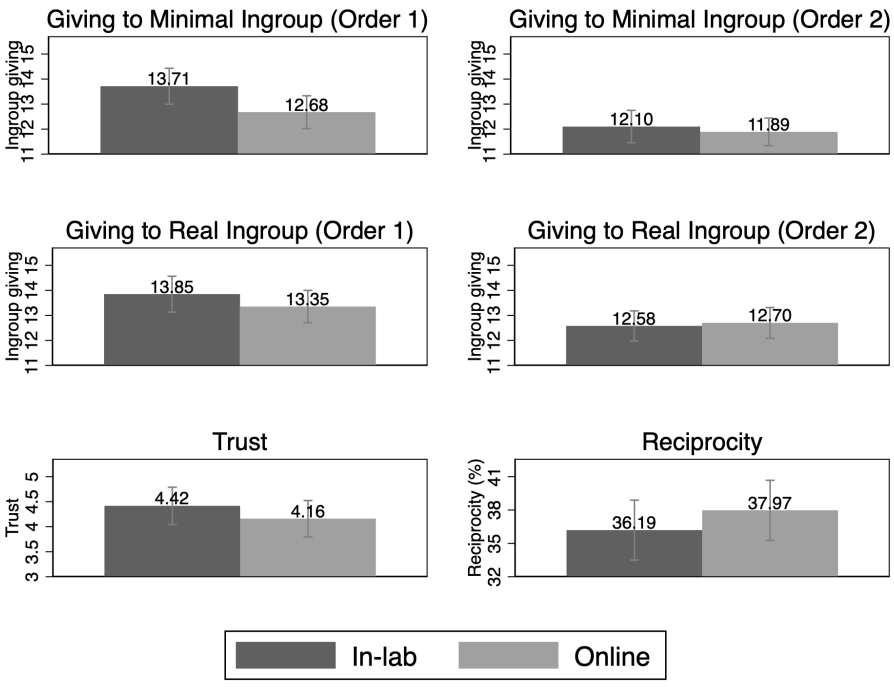
Table 1 details the means and p-values for the incentivized measures of social preferences. This table imposes a control for ordering effects for the dictator games. There is a difference between the lab and online subjects when the minimal group treatment is administered first. However, adjusting for multiple hypothesis testing, this difference disappears.

Table 1

	In-lab	Online	Difference	Unadj.p-val	Adj. p-val
Giving to Minimal Ingroup					
Order 1	13.712	12.676	1.036	0.037	0.178
	(0.365)	(0.333)	(0.494)		
	n =118	n =139			
Order 2	12.102	11.890	0.211	0.623	0.849
	(0.329)	(0.280)	(0.429)		
	n =118	n =146			
Giving to Real Ingroup					
Order 1	13.847	13.353	0.495	0.313	0.824
	(0.364)	(0.329)	(0.489)		
	n =118	n =139			
Order 2	12.576	12.699	-0.122	0.783	0.849
	(0.307)	(0.312)	(0.444)		
	n =118	n =146			
Trust	4.415	4.159	0.256	0.337	0.824
	(0.190)	(0.185)	(0.267)		
	n =236	n =283			
Reciprocity	36.193	37.965	-1.772	0.366	0.824
	(1.371)	(1.376)	(1.959)		
	n =236	n =283			

Notes: Standard errors in parentheses. Adjusted p-values are calculated using Stata command *wyoung* (Jones, Molitor, and Reif 2019). Order 1 = minimal first and real second; order 2 = minimal second, real first

Figure 25 graphs the mean for each incentivized measure and includes the 95 percent confidence interval. This figure illustrates what is detailed in Table 1. There is in-group bias in the dictator games and the effect is true for both laboratory and online subjects. There are no differences for the trust game.



Note: Error bars represent 95% confidence interval.

Figure 25

4 Recruitment Emails

The following are copies of the recruitment emails sent to subjects.

Michelle Harris (notification):

The Behavioral Research Lab (BRL) invites you to participate in a decision making study. The study will last less than 60 minutes and you will be paid for your participation. You will need to come to the BRL to participate. You are eligible to participate at a time and date of your own choosing.

To see what times are available please go to: http://brl.rice.edu/Study2020 . There you will be able to pick a time to participate.

Play games! Earn money! Help science!
This research study has been reviewed and approved by Rice University Institutional Review Board. If you have concerns regarding this study or questions regarding your rights as a study participant, please contact Stephanie Thomas, Compliance Administrator -IRB, at Rice University.
Email: irb@rice.edu or Telephone: 713-348-3586

Rick K. Wilson, Director BRL and Department of Political Science

Figure 26: Copy of email sent to subjects asking them to signup for an in-lab experiment.

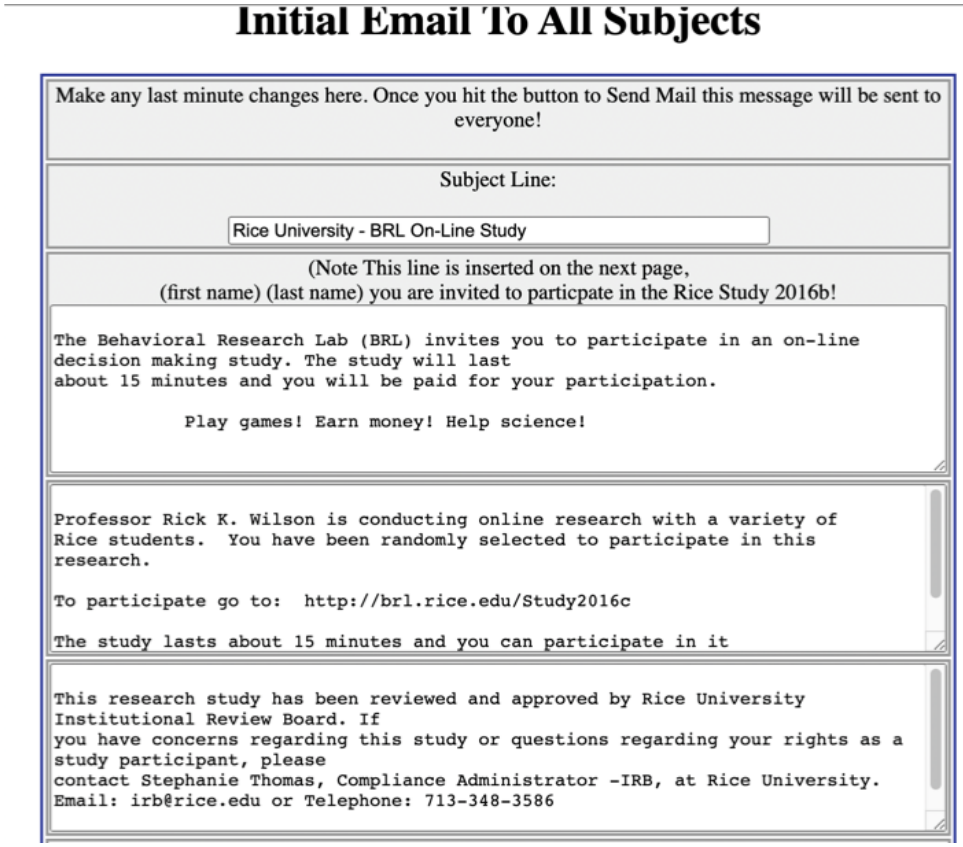


Figure 27: Copy of email sent to subjects asking them to begin an online experiment.