



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

Catherine Eckel¹ · Rick K. Wilson² · Sora Youn¹

Received: 1 August 2023 / Revised: 26 February 2024 / Accepted: 21 August 2024 /
Published online: 29 September 2024
© The Author(s), under exclusive licence to Economic Science Association 2024

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

1 Introduction

When COVID-19 disrupted laboratory experiments around the world, many researchers turned to online experiments. While the high level of control provided by the laboratory is considered the gold standard for conducting experiments with

✉ Rick K. Wilson
rkw@rice.edu

Catherine Eckel
ceckel@tamu.edu

Sora Youn
sorayoun@tamu.edu

¹ Department of Economics, Texas A&M University, College Station, TX 77845, USA

² Department of Political Science, Rice University, Houston, TX 77005, USA

human subjects, circumstances often lead to relaxing that control. Now that the pandemic is less of a threat scholars have the luxury of choice between engaging subjects online or returning to the lab. Subjects, too, have the choice to participate in laboratory experiments or opt only for experiments carried out online. Does it matter for experimental results if subjects self-select into their venue?

When measuring social preferences, this question is especially important. The lab may attract more socially motivated subjects. In addition, the lab facilitates closer monitoring of subjects and the lab diminishes potential doubts about the existence of counterparts participating in the experiment. This may lead to a finding of more generous social preferences in the lab when compared to those participating online. On the other hand, those who participate online do not have to make a significant time commitment by going to a lab. Those subjects might feel less deserving of compensation and translate their time savings into more generous behavior. As researchers, how much should we worry about the impact on our results of subjects who choose to participate in lab or in online settings?

In this study we invited all subjects who had participated in the first wave of a longitudinal panel to participate in a study measuring social preferences in our lab, located on campus. Those who signed up (the eager beavers) completed the task in a traditional laboratory setting. However, many of our participants had not signed up for this wave of the study. 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 to selection and to differences in the experience of in-lab compared to online participation. To our surprise, we find no significant differences in behavior between the two groups of subjects. This result should provide some reassurance to researchers when making inferences from subjects in one setting or the other.

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 that explicitly compare subjects recruited to the two venues conclude that online experiments are comparable to lab experiments (Horton et al. 2011; Brañas-Garza et al. 2018, 2023). We stress-test such results by asking whether this equivalence holds when subjects self-select into lab or online experiments.

To expand a bit, when studying social preferences, evidence from prior studies generally supports 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 three 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. Prissè 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 somewhat 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. We ask, if they are able to select when and where they participate, does that affect the findings?¹

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 the initial wave 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 (details concerning the sample are in the Supporting Information (SI) Sect. 1).

In Phase 2 all panelists were sent recruitment emails inviting them to sign up for an in-lab study. From November 2, 2016, to November 23, 2016, a total of 22 lab sessions were conducted with the 236 subjects who responded to our request. The remaining panelists who chose not to come to the lab were then recontacted and recruited for an online study. The online study was open from November 28, 2016 to April 28, 2017.² This yielded another 285 subjects. Note that subjects were *not* randomly assigned to one form of participation or the other. The "eager beavers" chose to come to the lab in response to the general call for subjects. The "lazy slugs" chose not to come to the lab, but later participated online. During the period of active recruitment into the lab, subjects did not know there would be a later opportunity to participate online. Subjects were already familiar with an online study, since they all had participated in Phase 1 prior to matriculating at Rice.

The study interface was *identical* for both the lab and online participants (see the SI Sect. 2). We used standard lab and online procedures. Subjects attending the lab

¹ Other studies are concerned about selection biases among subjects who choose to participate in experiments compared to the population from which they were recruited (e.g., Cleave et al. 2013). Our focus is with possible selection effects between subjects who participate in the lab and those who participate online.

² The online portion of the study was initiated at the end of the academic term. Over 93 percent of the subjects completed the online study by February 1, 2017. SI Sect. 4 reproduces the emails used for recruitment.

sessions read instructions and were able to ask questions, and completed their survey questions and tasks independently. Subjects in the online version read instructions and made decisions independently, on their own time, and were not observed.

In this paper, we focus on two third-party dictator games and a standard trust game (details are in the SI—Sect. 2).³ In the third-party dictator games, the dictator 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 making the decision. The two third-party dictator games differed in the definition of the groups. In one, the recipients were either from their own residential college (the in-group) relative to someone who was in a different residential college (a *real* group). In the other, we created artificial minimal groups, and recipients were part of the dictator's *minimal* in-group and someone who was in the out-group. Subjects completed the two games in random order, and we control for order effects in our discussion (for further details about these treatments, see our paper comparing the group types 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-preference 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 each possible amount that could be sent.

All subjects completed all of the tasks, and were informed that their randomly-assigned counterparts were also participants in the study. One of the tasks was randomly selected for payment for each subject. 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, with the dictator earning the \$5 fee. For the trust game, subjects played both roles, truster and trustee in that order, and at the end of the experiment were randomly assigned to one role. Thus both games have a "role uncertainty" design (Iriberry and Rey-Biel 2011). Subjects were not told which task was paid until the end of the experiment. All of the randomization and matching to positions was computerized. For subjects in the lab, matching was within session and subjects were paid their earnings at that point, as is customary for lab experiments. For subjects who were online, matching was with others participating online and payments were delayed until the end of the study. Subjects had previously experienced this form of matching and payment, which was used in the first wave. Subjects spent an average of 39 min in the lab (and less time online) and earned an average of \$21.37.

³ There are a number of experimental games designed to measure different aspects of social preferences. Because components of our study were designed for different purposes, we include here only two: third party dictator games, and trust games. Several studies have noted that social preferences are strongly positively correlated across measures. See for example, Peysakhovich and Nowak (2014).

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 subjects opting for the laboratory and those who subsequently took up the online option. Table 1 reports subject characteristics. We find that the eager beavers are very similar to the lazy slugs. There is balance between male and female subjects across venues. Asians are more likely to show up in the lab than Caucasians. When we look at a measure of risk aversion collected in the first-wave study (prior to matriculation), we find no difference between the two sets of participants. The same is true for a measure of time preferences.⁴ There are no differences in GPA measured in several ways. Across the five personality inventory items, we find no significant differences. As proxies for students' family wealth, we use the zip codes recorded at the time of subjects' entry to Rice University. *Median house value* represents the median housing costs, and *Median household income* indicates the median income by zip code. We find no significant differences in these measures.⁵ The differences we note lose statistical significance when adjustments are made for multiple hypothesis testing Westfall and Young (1993). All-in-all the two groups are well-balanced across multiple characteristics and measures.⁶

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 in-group member (relative to a non-in-group-member) from the \$20 budget. Both rows show in-group favoritism (amounts greater than \$10, more than half of the budget, are sent to the in-group members). Adjusting for multiple hypothesis testing leads to no significant differences being detected.⁷ The last two rows indicate that there are no significant differences in trust or reciprocity.⁸

Figures 1 and 2 graph the mean for each incentivized measure and includes the 95 percent confidence interval. This figure illustrates what is detailed in Table 2.

⁴ A discussion of these measures is reported in Eckel et al. (2023).

⁵ 67 students from outside the US are not included.

⁶ Table 1 of the Supplementary Information (SI) also presents adjusted p -values derived from statistics suggested by Bonferroni-Holm, Sidak-Holm, and Romano-Wolff. Notably, all of these approaches yield the same result.

⁷ Table 2 of the Supplementary Information (SI) also presents adjusted p -values derived from Bonferroni-Holm, Sidak-Holm, and Romano-Wolff techniques. Notably, all of these techniques yield the same result. We are also concerned with ordering effects for the dictator games. Table 3 of the Supplementary Information (SI) shows there is no difference when adjusting for multiple hypothesis testing.

⁸ In the online study, two observations were lost due to a programming error. Hence the differences in number of participants 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-value	Adj. p-value
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	0.17
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	0.96
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	0.99
GPA in the first semester	3.61	0.03	3.55	0.03	0.06	0.04	0.18	0.94
Cumulative GPA by Fall 2021	3.62	0.02	3.59	0.02	0.03	0.04	0.37	0.99
Number of completed hours by Fall	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	0.96
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	0.32
Openness to New Experiences	5.34	0.07	5.26	0.06	0.08	0.09	0.36	0.99
Wealth								
Median house value	315,758	15,814	343,953	15,631	-28,195	22,522	0.21	0.96
Median household income	84,869	2,396	84,430	2,205	439	3,270	0.89	1.00

Note: Westfall and Young (1993) adjusted *p*-values are estimated using 10,000 bootstraps. Stata command *wyouth* (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. To construct *Median house value* and *Median household income* as proxies for students' wealth level, we use the zip codes of all students at Rice University when they entered the institution. *Median house value* represents the median housing costs, and *Median household income* indicates the median income by zip code

Table 2 Means for Social Preference Measures

	In-lab	Online	Difference	Unadj. <i>p</i> -val	Adj. <i>p</i> -val
Giving to Minimal Ingroup	12.907 (0.251) n=236	12.274 (0.218) n=285	0.633 (0.331)	0.056	0.187
Giving to Real Ingroup	13.212 (0.241) n=236	13.018 (0.227) n=285	0.194 (0.332)	0.559	0.676
Trust	4.415 (0.190) n=236	4.159 (0.185) n=283	0.256 (0.267)	0.337	0.676
Reciprocity	36.193 (1.371) n=236	37.965 (1.376) n=283	-1.772 (1.959)	0.366	0.676

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

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.

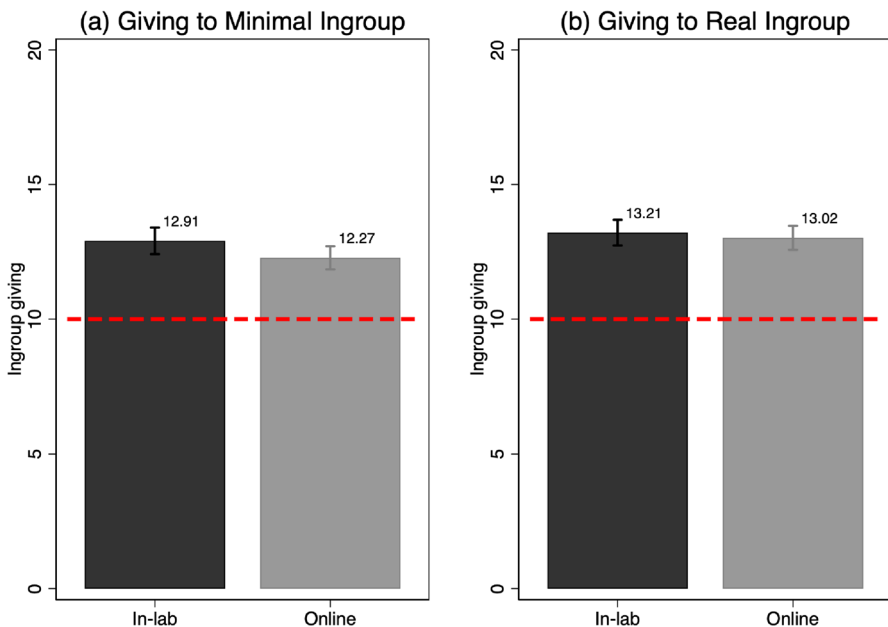


Fig. 1 Means for social preference measures: third-party dictator games. Note: The dashed line indicates an equal split between the ingroup and the outgroup. Standard error bars are included

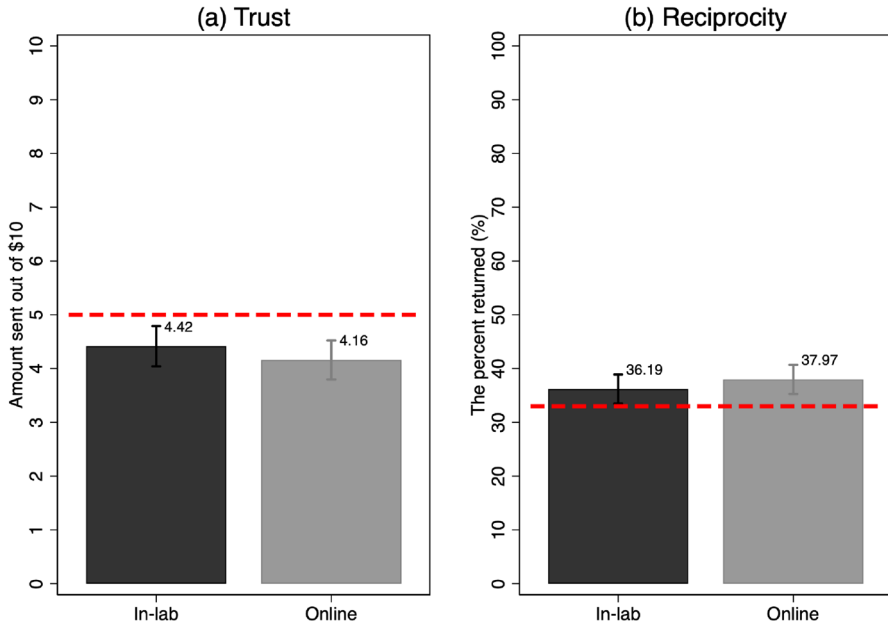


Fig. 2 Means for social preference measures: trust games. Note: The dashed line for trust indicates half the endowment was sent. The dashed line for reciprocity indicates the break-even point for whether trust paid. Standard error bars are included

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.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40881-024-00179-1>.

Acknowledgements 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.

Data availability Protocols, data, and computer code are available at: <https://doi.org/10.17605/OSF.IO/WD6UM>. Pre-registered hypotheses are available at: <https://doi.org/10.17605/OSF.IO/PZ5WG>.

Declarations

Conflict of interest The authors declare they have no conflict of interest.

References

- Akerlof, G. (1997). Social distance and social decisions. *Econometrica*, *65*(5), 1005–1028.
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142.
- Brañas-Garza, P., Capraro, V., & Rascon-Ramirez, E. (2018). Gender differences in altruism on mechanical turk: Expectations and actual behaviour. *Economics Letters*, *170*, 19–23.
- Brañas-Garza, P., Jorrat, D., Espín, A. M., & Sánchez, A. (2023). Paid and hypothetical time preferences are the same: Lab, field and online evidence. *Experimental Economics*, *26*(2), 412–434.
- Buso, I. M., Cagno, D. D., Ferrari, L., Larocca, V., Loré, L., Marazzi, F., Panaccione, L., & Spadoni, L. (2021). Lab-like findings from online experiments. *Journal of the Economic Science Association*, *7*(2), 184–193. <https://doi.org/10.1007/s40881-021-00114-7>
- Cleave, B., Nikiforakis, N., & Slonim, R. (2013). Is there selection bias in laboratory experiments? The case of social and risk preferences. *Experimental Economics*, *16*(3), 372–382. <https://doi.org/10.1007/s10683-012-9342-8>
- Clifford, S., & Jerit, J. (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., Shultz, T., & Onishi, K. (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., Sinha, N., & Wilson, R. (2023). The evolution of preferences and charitable giving: A panel study of the university years. *Oxford Economic Papers*, *2*, 1–21. <https://doi.org/10.1093/oeqf/gpad030>
- Eckel, C., & Wilson, R. (2006). Internet cautions: Experimental games with internet partners. *Experimental Economics*, *9*, 53–66. <https://doi.org/10.1007/s10683-006-4307-4>
- Eckel, C., Wilson, R., & Youn, S. (2022). In-group favoritism in natural and minimal groups. *Economics Letters*, *219*, 110794.
- Hergueux, J., & Jacquemet, N. (2015). 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., Rand, D., & Zeckhauser, R. (2011). 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>
- Iriberry, N., & Rey-Biel, P. (2011). The role of role uncertainty in modified dictator games. *Experimental Economics*, *14*, 160–180.
- Jones, D., Molitor, D., & Reif, J. (2019). 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/qjz023https://academic.oup.com/qje/article-pdf/134/4/1747/32666232/qjz023.pdf>
- Peysakhovich, A., & Nowak, M. (2014). 09. Humans display a “cooperative phenotype” that is domain general and temporally stable. *Nature Communications*, *5*, 4939. <https://doi.org/10.1038/ncomms5939>
- Prissè, B., & Jorrat, D. (2022). Lab vs online experiments: No differences. *Journal of Behavioral and Experimental Economics*, *100*, 101910. <https://doi.org/10.1016/j.socce.2022.101910>
- Westfall, P. H., & Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*. John Wiley & Sons.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.