ISSN (Print) 0473-453X
Discussion Paper No. 1168
ISSN (Online) 2435-0982

COMPARING DATA GATHERED IN AN ONLINE AND A LABORATORY EXPERIMENT USING THE TRUSTLAB PLATFORM

Nobuyuki Hanaki
Takahiro Hoshino
Kohei Kubota
Fabrice Murtin
Masao Ogaki
Fumio Ohtake
Naoko Okuyama

Revised June 2022
March 2022

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

Comparing data gathered in an online and a laboratory experiment using the Trustlab platform*

Nobuyuki Hanaki† Takahiro Hoshino‡ Kohei Kubota§
Fabrice Murtin¶ Masao Ogaki‖ Fumio Ohtake **
Naoko Okuyama††

June 4, 2022

Abstract

This paper compares the results of an experiment conducted both in the laboratory and online with participants recruited from the same subject pool using the Trustlab platform. This platform has been used to obtain incentivized and internationally comparable behavioral economics measures of altruism, cooperation, reciprocity, trust, and trustworthiness, employing representative samples in many countries.

*The experiment reported in this paper was approved by the Institutional Review Board of the Institute of Social and Economic Research, Osaka University. We gratefully acknowledge that this work was partially supported by the Japan Society for the Promotion of Science KAKENHI Grant nos. 19H00599 and 21K18129, the Joint Usage/Research Center at the Institute of Social and Economic Research, Osaka University, and the Keio University Academic Development Fund.

†Institute of Social and Economic Research, Osaka University. E-mail: nobuyuki.hanaki@iser.osaka-u.ac.jp

‡Faculty of Economics, Keio University. E-mail: hoshino@econ.keio.ac.jp

§Faculty of Commerce, Chuo University. E-mail: kkubota@tamacc.chuo-u.ac.jp

¶OECD Centre on Well-Being, Inclusion, Sustainability and Equal Opportunity (WISE) E-mail: Fabrice.MURTIN@oecd.org

‖Corresponding author. Faculty of Economics, Keio University. E-mail: mogaki@econ.keio.ac.jp

**Center for Infectious Disease Education and Research, Osaka University E-mail: ohtake@econ.osaka-u.ac.jp

††Graduate School of Environment and Information Sciences, Yokohama National University. E-mail: okuyama-naoko-rt@ynu.ac.jp

We find little significant difference between the results from sessions conducted in the laboratory and online. While the existing literature shows that the choice between laboratory and online experiments can cause differences in results in some cases, our findings support the hypothesis that they do not cause differences in the behavioral economics measures when using the Trustlab platform.

Keywords: dictator game, trust game, public goods game
JEL Code: C90

1 Introduction

The current COVID-19 pandemic has promoted many changes in society, with many activities, including business meetings and university lectures, now done online. Research employing the method of experimental economics is no exception, with many researchers, thanks to the development of tools such as Chen et al. (2016) and Duch et al. (2020), now conducting experiments, hitherto done only in a laboratory prior to the pandemic, online.

Another important reason for conducting online experiments is that it is less costly to conduct artefactual field experiments (in the terminology of HarrisonList2004 with a nonstandard subject pool (say, a national representative pool rather than the standard students subject pool in the conventional laboratory experiment).

A natural concern of running experiments online instead of in a laboratory is whether doing so will have a significant impact on the results. We may be particularly concerned about researchers losing control in terms of the information to which participants have access and participants being more easily distracted during online experiments compared with in-laboratory experiments. To address this concern, we report the results of an identical

experiment conducted in both laboratory and online settings with participants recruited from the same subject pool.

For the same subject pool, we use the student sample. Just as LevittList2007 argued that a field experiment can serve as a bridge connecting lab-generated data and data from natural settings, a student sample online experiment can serve as a bridge connecting lab-generated data and data from natural settings.

Few extant studies compare the behavior observed in experiments conducted online and in a laboratory using participants recruited from the same subject pool. The results are already mixed, as we explain in detail in the following section. Some find no significant differences between online and laboratory sessions, while others expose significant differences. We suggest this shows that online and laboratory experiments can provide differences in results, but that this depends on the experiment and the platform used to conduct it. Therefore, it is important to examine whether online and laboratory experiments account for differences in results for different experiments and dissimilar platforms.

This paper focuses on behavioral economics measures of altruism, cooperation, reciprocity, trust, and trustworthiness in dictator, public goods, and trust game experiments using the Japanese version of the OECD's Trustlab project platform (see, Murtin et al., 2018).[1] The Trustlab platform has

[1]These measures have been used in many studies to examine cooperative behavior (see, e.g., Fischbacher et al., 2001; Romano et al., 2017; Algan, 2018). For example, Falk et al. (2018) included some of these in a questionnaire administered in several countries after verifying that the results would be consistent with comparable laboratory experiments. As Ogaki (2022) argues, these measures are important for understanding the community mechanism, whereby older people have greater difficulty in effectively using the market mechanism alone as their cognitive ability declines with the natural aging process or due

already been used to collect data on these behavioral economics measures through incentivized online experiments together with other data, such as preferences for redistribution, trust in government, and attitudes towards migrants, through non-incentivized questionnaires.[2]

In the main Trustlab Project, a representative data set (in the sense of representativeness for sex, age, and income) for 1,000 or more individuals is collected in eight countries (France, Germany, Italy, Japan, Korea, Slovenia, UK, and US). Murtin et al. (2021) provide evidence that these measures help explain key political preferences, such as preferences for redistribution, trust in government, and attitudes towards migrants, in these eight countries.[3]

Some authors of this paper are core members of the main Trustlab project in Japan. Our Trustlab Japan project extends the main projects of Trustlab in two dimensions. First, instead of the cross-sectional data in the main project, we collect panel data for three waves to investigate the dynamics of these behavioral measures.[4] Second, we use the Trustlab platform to conduct online experiments with student groups to address potential concerns regarding external validity, in particular, "the representativeness of the sampled populations" (List, 2007), for various experiments conducted at Osaka

to dementia.

[2]Although the experiments involve interactive decisions in trust game and public goods games, because a strategy method is used (as described in detail below), each participant could complete the experiment individually.

[3]Aassve et al. (2018) present data and the peculiarities of Trustlab Italy, and Cetre et al. (2020) study ethic in-group bias in trust games in a module on ethnic discrimination using Trustlab projects in the US and Germany.

[4]For this reason, we begin the first wave with a sample of about 2,500 participants, so that we can obtain panel data for a sample of about 1,000 participants after some drop out in the following two waves. At the time of writing, the Trustlab panel data in the main project has only been collected in Japan. We completed data collection for three waves: the first wave in January–February 2020 with a sample size of 2,504, the second wave in June–July 2020 with a sample size of 1,520, and a third wave in September–October 2021.

University (OU), a highly selective university in Japan.

At this point, we have only preliminary results for the panel data.[5] This paper focuses on the student sample for the purpose of comparing online and laboratory experiments. Ours is the first study to compare these with the Trustlab platform in any country. In Appendix B, we compare our results from the OU sample and those from our representative sample in the first wave with the largest sample size.

Our data reveals that there is little significant difference between the results from sessions conducted in the laboratory and online for four incentivized tasks. This supports our hypothesis that laboratory and online experiments do not cause differences in results for the behavioral economics measures to these same tasks in the Trustlab platform.

2 Literature review

As discussed, only a few existing studies compare the behavior observed in experiments conducted online and in a laboratory using participants recruited from the same subject pool. The results are mixed. Some find no significant differences between the online and the laboratory sessions (Hergueux and Jacquemet, 2015; Ozono and Nakama, 2021; Snowberg and Yariv, 2021), while others find significant differences (Prissé and Jorrat, 2021; Schmelz and Ziegelmeyer, 2020).

For example, Hergueux and Jacquemet (2015) measure social preference

[5]The preliminary results show that incentivized measures of trust, altruism, positive reciprocity, and cooperation move together for many participants between the first two waves. There is some evidence that part of these comovements for women was caused by COVID-19.

both online and in a laboratory, based on a public goods game, a dictator game, an ultimatum bargaining game, and a trust game, together with an elicitation of risk aversion. While finding no significant differences between the two samples, the study noted that contrary to the prediction of social distance theory, more other-regarding decisions are observed in the online sample.[6]

Prissé and Jorrat (2021) compare online and laboratory sessions concerning time and risk preferences, the degree of altruism within the dictator game in which subjects decide whether to donate their total earnings, and cognitive ability.[7] They also report no significant difference between the two samples except in the dictator game, even with significantly more participants in the online than laboratory sessions. This contrasts with the findings in Hergueux and Jacquemet (2015).

Ozono and Nakama (2021) compare the results of online and laboratory experiments in repeated public goods games with and without punishment, as well as for individual tasks. Ozono and Nakama (2021) report no significant difference between the laboratory and the online data except for a creative individual task with an external incentive. For this task, performance was lower for the laboratory session than the online session.[8]

[6]Hergueux and Jacquemet (2015) also provide a survey of earlier studies.

[7]Time preference is measured by the convex time budget task Andreoni and Sprenger (2012) and a modified version of the multiple price list of Andreoni et al. (2015). Risk preference is measured by the multiple price list of Holt and Laury (2002). Cognitive ability is measured using the cognitive reflection test in Frederick (2005) and a numeracy task.

[8]Ozono and Nakama (2021) also compare the data gathered for the same set of tasks between a student sample and a sample of online workers recruited from Yahoo crowdsourcing, and report that their sample of online workers contributes, on average, significantly less in the public goods game than their student sample.

Schmelz and Ziegelmeyer (2020) investigate the effect of managerial control on worker's (costly) effort in a principal–agent game in laboratory and online experiments. They find that while efforts increase with the level of control in both the laboratory and online sessions, the effect of control is significantly stronger in the online experiment. Moreover, in the absence of control, workers' effort level is significantly lower in the online sessions than the laboratory session. They also find that reciprocity (by workers) is significantly weaker in the online than the laboratory sessions.

Snowberg and Yariv (2021) report, among other things,[9] the results of a comparison between an online and laboratory experiment where the same set of participants participated in both only a few months apart. They find no significant difference in the elicited measures except for students being more risk averse in the laboratory than online and performing better in two cognitive tasks in the laboratory over online.

3 Experiment

The experiments are conducted using the platform for the Japanese version of the OECD's Trustlab project (see Murtin et al., 2018), which is programmed using oTree (Chen et al., 2016).

The platform contains four tasks with monetary incentives as well as a non-incentivized questionnaire. The four incentivized tasks are for the trust

[9]Snowberg and Yariv (2021) investigate the possible effects of students self-selecting into laboratory experiments by comparing students from a representative US sample with a sample of MTurkers. Using data gathered via a university-wide survey conducted at CalTech (thus with little self-selection), they find only slight differences between those that participate and those that do not in laboratory experiments.

game, the public goods game, the dictator game, and a risky lottery choice, implemented in order before the non-incentivized questionnaire. The game tasks employ the strategy described below.

In the trust game, a participant is randomly paired with another participant. One is chosen randomly as the first mover, and another as the second mover. Each participant receives an endowment of 1,000 JPY. The first mover then decides to send an amount $X \in [0, 1000]$ from their endowment, which will be tripled by the experimenter, to the second mover. The second mover then decides to send back an amount $Y \in [0, 1000 + 3X]$ to the first mover. Participants are then asked (a) how much, out of the initial 1,000 JPY endowment, would they send to their randomly chosen partner if they are chosen as the first mover, as well as (b) how much they would return, conditional on the each of the 11 amounts, $\in \{0, 100, 200, \ldots, 900, 1, 000\}$ JPY, sent by the first mover, to the first mover when chosen as the second mover. Participants were also asked how much they expect a randomly chosen second mover, when they as the first mover have sent 500 JPY, would send back (out of 2,500 JPY). This question, however, is not incentivized.

In the public goods game, a group of four participants is randomly created. Each has a 1,000 JPY endowment. Each participant then decides to invest $I^i \in [0, 1000]$ in a project. The amount not invested is retained by each of the participants. The amount invested by the four participants is then multiplied by 1.6, and the resulting sum divided and returned equally to each participant, regardless of the amount invested. Participants are asked (a) how much they would invest in the project without knowing how much others in the group have invested, and (b) how much they would invest in the

project if the average investment of the others is $\{0, 100, 200, \ldots, 900, 1000\}$ JPY.

In the dictator game, each participant is randomly paired with another. Participants are then asked if they were chosen as the dictator, how much of a 1,000 JPY endowment they would give to a randomly paired participant. If chosen to be the recipient, they receive the amount given by the dictator. Finally, in the risky lottery choice, each participant is asked to choose from one of six lotteries: 1:(800, 800), 2:(700, 1000), 3:(600, 1200), 4:(500,1400), 5:(400, 1600), and 6:(100, 1900), where each lottery would pay one of the two amounts in parenthesis (both in JPY) with equal probability.

After the experiment, one of the four tasks is assigned randomly (conditional that the required number of participants can be assigned to form a group) to participants to compute their reward. Participants received, in addition to the participation fee of 500 JPY, the amount based on their own decisions as well as the decisions of others in their randomly formed group (or pair). The payment was in cash for the laboratory experiment and in the form of an Amazon Gift Card (e-mail version) for the online experiment. Participants were informed of the method of payment as well as the amount of the participation fee when registering for the experiment.

4 Result

Laboratory experiments were conducted between November 17 and 20, 2020. A total of 84 students participated in over four sessions.[10]

[10]The number of participants in each session was 24, 24, 20, and 16. These variations arise from differences in the number of registered participants across sessions. Participants

The online experiment was conducted on December 7, 2020. A total of 116 students participated. As with the laboratory experiments, participants were asked to register to participate in the experiment. On the day of the experiment, however, each participant received a customized link via e-mail and participated in the experiment individually by clicking the link.[11]

While the existing studies reviewed in Section 2 randomly allocate participants into online and laboratory sessions to control for participants self-selecting into one of the sessions, we have not done this in our experiment because we were also interested in testing if self-selection would also result in significant differences in the observed behavior.

Below, we report the results of the three tasks, namely, the dictator, trust, and public goods games, with monetary incentives. We employ a 5% significance level. The results of the risk preference elicitation are reported in Appendix A. Based on the data gathered in these three games, we construct measures of altruism, cooperation, reciprocity, trust, trustworthiness, and trustworthiness2. See Table 1 for the definition of each measure.

Figure 1 reports the distributions of, for both the online and laboratory samples (shown in red and blue, respectively), altruism, cooperation, reciprocity, trust, trustworthiness, and trustworthiness2. All reported are p-values for two-tailed Mann–Whitney (MW) and Kolmogorov–Smirnov (KS) tests. As shown, only the distribution of trust (but not the median) is significantly different between the two samples at the 5 % level (p=0.040 for

are recruited through the subject database managed by ORSEE (Greiner, 2015).

[11]Because the experiment consists only of individual tasks, we did not use a teleconferencing tool (such as Zoom) to first gather participants together to check participation and to provide instruction. The experiment for the Japanese representative sample was also conducted by sending a customized link via e-mail.

Altruism	Amount given in the dictator game
Cooperation	Amount invested in the public goods game unconditional on the average investment of others
Reciprocity	Average difference between the conditional amount invested in the public goods game and the average investment of others
Trust	Amount transferred as first mover in the trust game
Trustworthiness	Average amount transferred back as second mover in the trust game
Trustworthiness2	Average share of available resource transferred back as second mover in the trust game

Table 1: Definitions of the six measures

the KS test). This is a multiple test problem for a set of 6 hypotheses. With the Bonferooni correction nor the Holm correction, we do not reject the null hypothesis even for the KS test. Thus, we have little evidence for significant difference.

In Appendix A, we also provide a more detailed comparison of the data behind the measures of reciprocity and trustworthiness2, namely, the amount of contribution conditional on the average contribution of others in the public goods game and the share of available resources transferred back conditional on the amount given by the first mover in the trust game. As before, we do not observe any significant difference between the two samples.

5 Concluding remarks

With the current COVID-19 pandemic, many applied researchers, including ourselves, have begun conducting experiments online instead of in the laboratory. Because online experiments are new, some may be concerned about the possible impact of experimenters losing control on the information

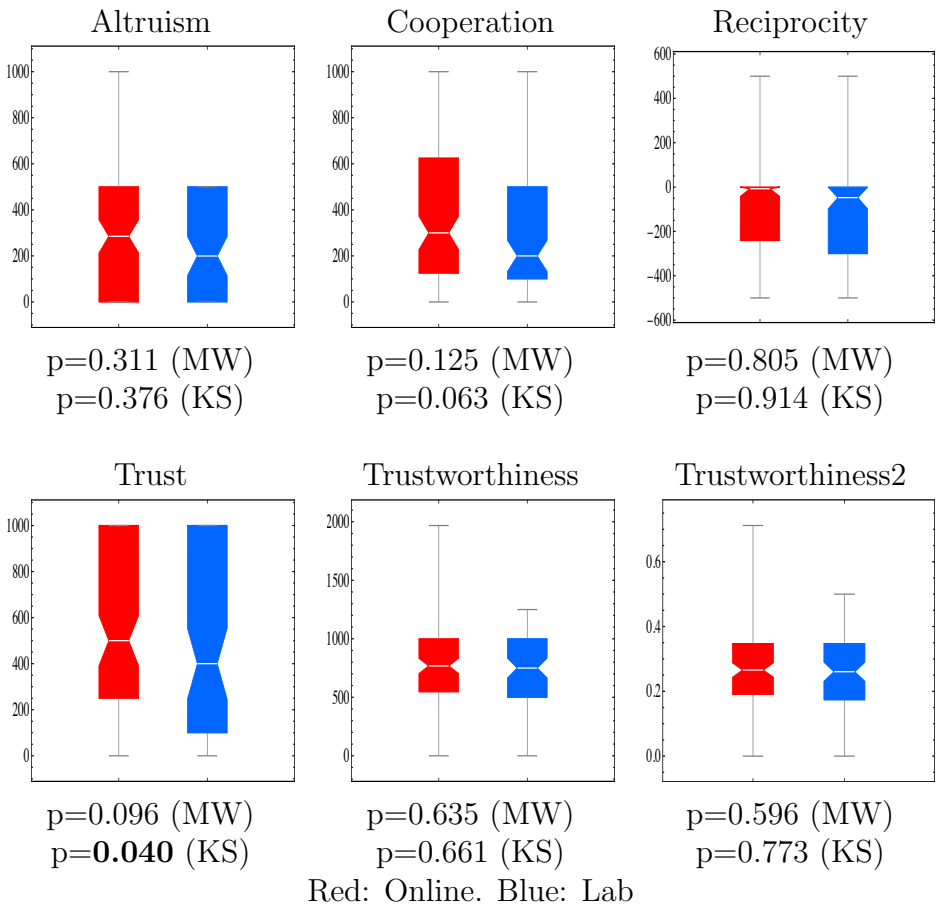


Figure 1: Distributions of behavioral measures

participants can access or participants being more easily distracted during online experiments than laboratory experiments. Some of the literature has addressed these concerns by reporting the results of experiment conducted in both the laboratory and online with participants recruited from the same subject pool. Because existing results are mixed, we believe that whether laboratory and online experiments provide differences in results depends on variation in the experimental tasks and changes in platforms.

To investigate this further, this paper focused on the behavioral economics

measures in trust, public goods, and dictator games using the Trustlab platform. Our data revealed that there is little significant difference between the results from sessions conducted in the laboratory and online. Thus, the abovementioned concerns do not appear to cause any major problems for these experiments when using the Trustlab platform.

Online experiments allow access to a wider set of participants, such as a representative sample of the populations of different countries, than laboratory experiments. Having established the absence of substantial significant differences between laboratory and online experiments, we are more confident in conducting online experiments to compare the results obtained from a student sample and a nonstudent sample to address the concerns related to external validity. In particular, the "representativeness of the sampled population" (List, 2007) of the laboratory experiment in Snowberg and Yariv (2021).

In Appendix B, we report a comparison between the data from OU and the first wave of TrustLab Japan based on a representative sample of the Japanese population (in terms of age, sex, and income) conducted from January to February 2020. The results are significantly different, except for the median amount sent as first mover in the trust game ("trust"). We find the representative sample, on average, suggests higher altruism (gives more in the dictator game), higher trustworthiness (returns more as the second mover in the trust game), and higher reciprocity and cooperation (cooperates more, both unconditionally and conditionally, in the public goods game), and takes less risk than the OU sample.

These results are in line with other studies comparing students and non-

student samples (see, for example, Cooper and Kagel, 2015; Fréchette, 2015; Snowberg and Yariv, 2021). They also complement other studies comparing individual characteristics, such as cognitive and emotional ability and personality traits, of a large sample of OU students and another representative sample of the Japanese adult population (Hanaki et al., 2022). However, we do not elaborate further here because the data for the representative sample was gathered prior to the COVID-19 pandemic. We leave a comparison of the OU student sample and the representative sample to future research.

References

Aassve, A., L. Mencarini, F. Chiocchio, F. Gandolfi, A. Gatta, and F. Mattioli (2018): "Trustlab Italy; a new dataset for the study of trust, family demography and personality," Working Paper 115, The Dondena Centre, Univesità Bocconi.

Algan, Y. (2018): "Trust and social capital," in *For good meeasure: Advancing reseearch on weell-being metrics beyond GDP*, ed. by J. E. Stigliz, J.-P. Fitoussi, and M. Durand, Paris: OECD Publishing, chap. 4, 283–320.

Andreoni, J., M. A. Kuhn, and C. Sprenger (2015): "Measuring time prefer- ences: A comparison of experimental methods." *Journal of Economic Behavior and Organization*, 116, 451–464.

Andreoni, J. and C. Sprenger (2012): "Estimating Time Preferences from Convex Budgets," *American Economic Review*, 102, 3333–3356.

Cetre, S., Y. Algan, G. Grimalda, F. Murtin, L. Putterman, U. Schmidt, and V. Siegerink (2020): "Ethnic bias, economic success and trust: Findings from large sample experiments in Germany and the United States through the Trustlab platform," Working Paper 2020/04, OECD Statistics.

Chen, D. L., M. Schonger, and C. Wickens (2016): "oTree – An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 9, 88–97.

Cooper, D. J. and J. H. Kagel (2015): "Other-Regarding Preferences: A Selective Survey of Experimental Results," in *Handbook of Experimental Economics*, ed. by J. H. Kagel and A. E. Roth, Princeton, NJ: Princeton University Press, vol. 2, chap. 4, 217–289.

Duch, M. L., M. R. P. Grossmann, and T. Lauer (2020): "z-Tree unleashed: A novel client-integrating architecture for conducting z-Tree experiments over the Internet," *Journal of Behavioral and Experimental Finance*, 28, 100400.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018): "Global evidence on economic preferences," *Quarterly Journal of Economics*, 133, 1645–1692.

Fischbacher, U., S. Gächter, and E. Fehr (2001): "Are people conditionally cooperative? Evidence from a public goods experiment," *Economics Letters*, 71, 397–404.

Fréchette, G. R. (2015): "Experimental economics across subject populations," in *Handbook of Experimental Economics*, ed. by J. H. Kagel and A. E. Roth, Princeton University Press, vol. 2, chap. 7, 435–480.

Frederick, S. (2005): "Cognitive reflection and decision making," *Journal of Economic Perspectives*, 19, 25–42.

Greiner, B. (2015): "An Online Recruitment System for Economic Experiments," *Journal of the Economic Science Association*, 1, 114–125.

Hanaki, N., K. Inukai, T. Masuda, and Y. Shimodaira (2022): "Comparing Behavior Between a Large Sample of Smart Students and a Representative Sample of Japanese Adults," ISER DP 1160, Institute of Social and Economic Research, Osaka University.

Hergueux, J. and N. Jacquemet (2015): "Social preferences in the online laboratory: a randomized experiment," *Experimental Economics*, 18, 251–283.

Holt, C. A. and S. K. Laury (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655.

List, J. A. (2007): "Field experiments: A bridge between lab and naturally occuring data," *The B.E. Journal of Economic Analysis & Policy*, 6, 000010220215380637174 7.

Murtin, F., J. Bonnet, A. Aasave, Y. Algan, R. Boarini, L. Fleischer, M. M. Galizzi, S. Gonzalez, G. Grimalda,

R. Hortala-Valive, S. Kim, N. Martin, M. Ogaki, L. Putterman, U. Schmidt, and V. Si (2021): "Trust and its determinants: Evidence from the Trustlab experiment," Manuscript, OECD.

Murtin, F., L. Fleischer, V. Slegerink, A. Aassve, Y. Algan, R. Boarini, S. González, Z. Lonti, G. Grimalda, R. H. Vallve, S. Kim, D. Lee, L. Putterman, and C. Smith (2018): "Trust and its determinants: Evidence from the Trustlab experiment," Working Paper 2018/02, OECD Statistics.

Ogaki, M. (2022): "Economics of the Community Mechanism," Working paper, Keio Univesity.

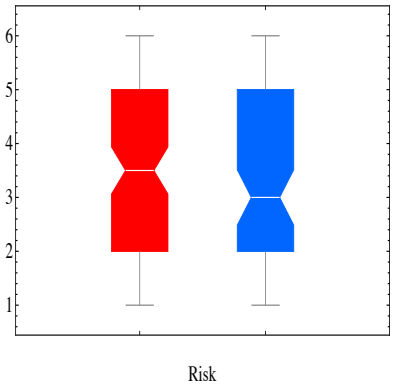
Ozono, H. and D. Nakama (2021): "Effects of experimental situation on group cooperation and individual performance: comparing laboratory and online experiments," Pre-print available at https://doi.org/10.31234/osf.io/847h2.

Prissé, B. and D. Jorrat (2021): "Lack of Control: An experiment," Tech. Rep. 3932079, SSRN.

Romano, A., D. Balliet, T. Yamagishi, and J. H. Liu (2017): "Parochial trust and cooperation across 17 societies," *Proceedings of the National Academy of Sciences, USA*, 114, 12702–12707.

Schmelz, K. and A. Ziegelmeyer (2020): "Reactions to (the absence of) control and workplace arrangements: experimental evidence from the internet and the laboratory," *Experimental Economics*, 23, 933–960.

Snowberg, E. and L. Yariv (2021): "Testing the Waters: Behavior across participant pools," *American Economic Review*, 111, 687–719.



p=0.569 (MW) p=0.716 (KS)
Red: Online. Blue: Lab

Figure A.1: Choice distributions in the risky lottery

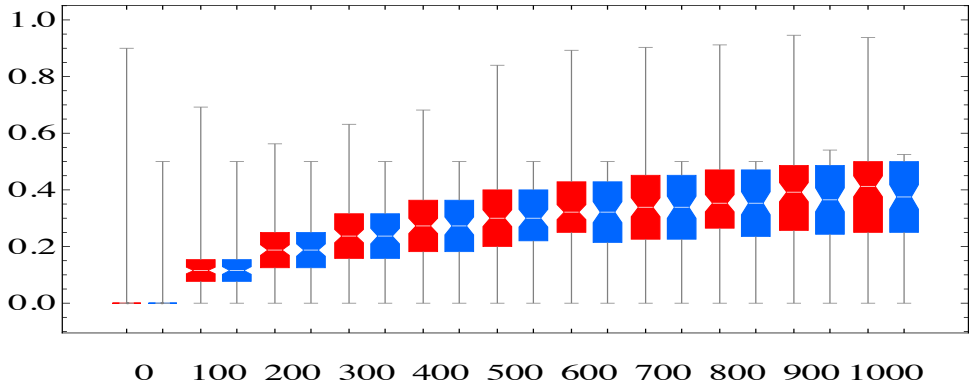
A Additional results

Figure A.1 plots the distributions of the choices in the risky lottery (shown in red and blue). There is no statistically significant difference between the two samples.

Panel (a) in Figure A.2 illustrates the distributions of the return rate in the trust game, conditional on the amount received, for the online and laboratory samples (shown in red and blue, respectively). The return rate is defined as the share of the amount sent back from the amount after receiving the transfer. For both samples, the return rates increase with the amount received, and there is no significant difference between the two samples.

Panel (b) in Figure A.2 depicts the distributions of the investment amount, conditional on the average investment by other members of the group, for the online and laboratory samples (shown in red and blue, respectively). Both samples demonstrate conditional cooperation in that the median investment matches the average investment by other members. There is no significant

(a) Distribution of return shares in the trust game

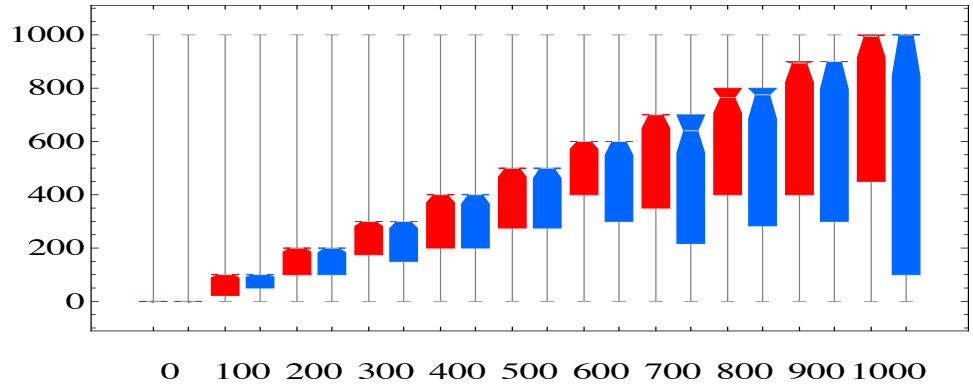


Red: Online. Blue: Lab

P-values of Mann–Whitney (MW) and Kolmogorov–Smirnov (KS) test

	0	100	200	300	400	500	600	700	800	900	1000
MW	0.058	0.526	0.596	0.827	0.963	0.770	0.767	0.572	0.374	0.578	0.314
KS	0.088	0.705	0.698	0.986	0.949	0.914	0.972	0.893	0.609	0.506	0.277

(b) Distribution of conditional cooperation in the public goods game



Red: Online. Blue: Lab

P-values of Mann–Whitney (MW) and Kolmogorov–Smirnov (KS) test

	0	100	200	300	400	500	600	700	800	900	1000
MW	0.591	0.837	0.594	0.796	0.869	0.960	0.810	0.749	0.894	0.958	0.311
KS	0.831	0.937	0.825	0.691	0.933	0.943	0.981	0.880	0.950	0.986	0.654

Figure A.2: Distributions of (a) return shares in trust game and (b) conditional cooperation in public goods game

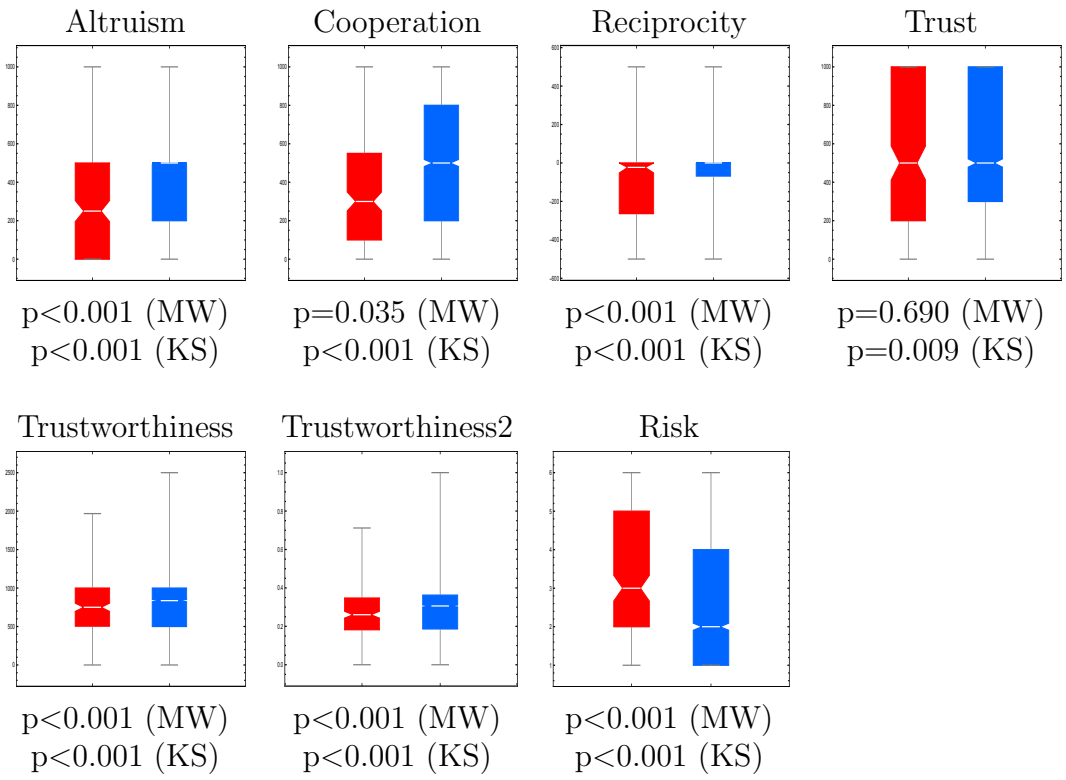


Figure B.1: Distribution of behavioral measures

difference between the two samples.

B Comparison of OU sample and non-OU sample

This Appendix reports the comparison between the results of the OU sample (pooling the online and laboratory samples) and the first wave of the TrustLab Japan questionnaire conducted online in January–February 2020.

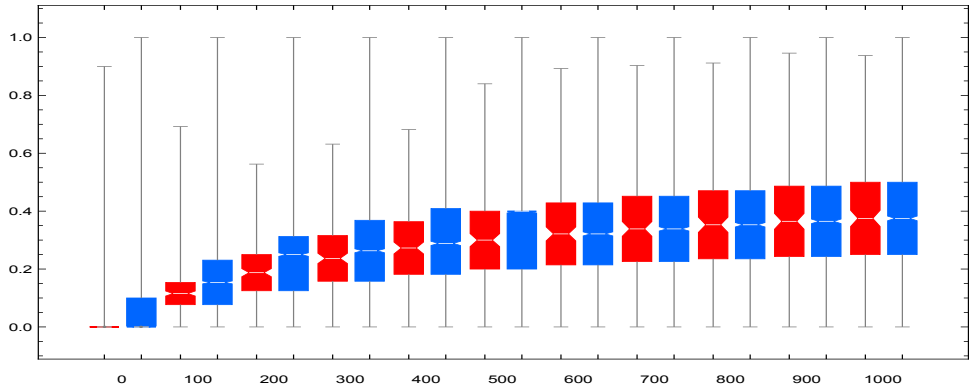
Figure B.1 reports the distributions of altruism, cooperation, reciprocity, trust, trustworthiness, and trustworthiness2 for the OU and non-OU samples

(shown in red and blue, respectively). Also reported are the p-values for the two-tailed Mann–Whitney (MW) and Kolmogorov–Smirnov (KS) tests. As shown, the results are significantly different, except for median trust, for the OU and non-OU samples. Namely, the representative sample, on average, shows higher altruism (gives more in the dictator game), higher trustworthiness (returns more as the second mover in the trust game), and higher reciprocity and cooperation (cooperates more, both unconditionally and conditionally, in the public goods game), and takes less risk than the sample from OU.

Panel (a) in Figure B.2 plots the distributions of the return rate in the trust game, conditional on the amount received, for the OU and non-OU samples (shown in red and blue, respectively). For both samples, the return rates increase with the amount received, but the rate is significantly higher for the non-OU sample than the OU sample.

Panel (b) in Figure B.2 depicts the distributions of the investment amount, conditional on the average investment by other members of the group, for the OU and non-OU samples (shown in red and blue, respectively). Both samples demonstrate conditional cooperation in that the median investments match the average investment by other members. There is a significant difference between the two samples. In particular, the interquartile range (25%–75%) is much narrower for the non-OU sample.

(a) Return shares in the trust game

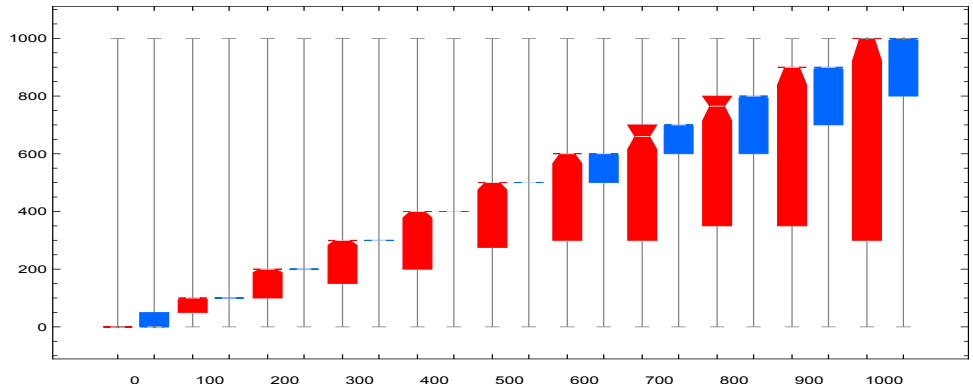


Red: OU sample. Blue: non-OU sample

P-values of Mann–Whitney (MW) and Kolmogorov–Smirnov (KS) tests

	0	100	200	300	400	500	600	700	800	900	1000
MW	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.002
KS	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.005	0.001	0.009	0.010	0.018

(b) Conditional cooperation in the public goods game



Red: OU sample. Blue: non-OU sample

P-values of Mann–Whitney (MW) and Kolmogorov–Smirnov (KS) tests

	0	100	200	300	400	500	600	700	800	900	1000
MW	<0.001	<0.001	<0.001	<0.001	0.001	0.001	0.002	<0.001	0.001	0.002	0.006
KS	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	0.002	0.006

Figure B.2: Distribution of (a) return shares in the trust game and (b) conditional cooperation in the public goods game