Growth and inequality in public good provision | an extended replication

Hauke Roggenkamp[a,b]

[a] *Helmut Schmidt University, Holstenhofweg 85, Hamburg, 22043,*

[b] *University of St. Gallen, Torstrasse 25, St. Gallen, 9000,*

Abstract

This is the abstract. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum augue turpis, dictum non malesuada a, volutpat eget velit. Nam placerat turpis purus, eu tristique ex tincidunt et. Mauris sed augue eget turpis ultrices tincidunt. Sed et mi in leo porta egestas. Aliquam non laoreet velit. Nunc quis ex vitae eros aliquet auctor nec ac libero. Duis laoreet sapien eu mi luctus, in bibendum leo molestie. Sed hendrerit diam diam, ac dapibus nisl volutpat vitae. Aliquam bibendum varius libero, eu efficitur justo rutrum at. Sed at tempus elit.

Keywords: Replication study, Non-convenience sample, Open science, Dynamic public good game, Online experiment, Generalizability

1. Introduction

> "There are two possible articles you can write: (a) the article you planned to write when you designed your study or (b) the article that makes the most sense now that you have seen the results. They are rarely the same, and the correct answer is (b)." (Bem, 1987, p. 171)

2. Methodology

In the terminology of Hamermesh (2007), I ran both a *pure* as well as a *scientific replication* of one treatment arm of Gächter et al. (2017)'s dynamic public good game. The pure replication re-analyzes the original data. Appendix A documents errors I identified in the original paper. The scientific replication, where I utilize a different sample drawn from a different population in a different situation, is described in this section.

2.1. Procedure

Participants entered the experiment at appointed times remotely from home. They first saw a welcome screen. After agreeing to the privacy policy, they could proceed to the instructions individually. Having read these instructions, each participant has also seen a demo-screen explaining the user interface. Before proceeding, they had to answer six comprehension questions correctly. Subsequently, they saw a waiting screen until they could be matched with three other participants, who have answered the comprehension questions correctly. Once matched, they were exposed to the decision screen over ten periods. At the end of the last period, participants saw results of all periods. Subsequently, they were exposed to a voluntary climate action, where they could donate (some of) their earnings to offset carbon dioxide. Subsequently, I elicited risk preferences (Holt and Laury, 2002) and finished with Gächter et al. (2017)'s questionnaire.

*Corresponding author
Email address: hauke.roggenkamp@unisg.ch (Hauke Roggenkamp)

2.2. Experimental Design

As in the NOPUNISH 10 Period treatment arm of Gächter et al. (2017), I ran sessions with groups of four ($i \in I = \{1, 2, 3, 4\}$), an initial endowment of $N_i^1 = 20$ tokens, $T = 10$ rounds, a private account with a return of 1 and a group account with a return of 1.5 ($\Rightarrow$ MPCR$\equiv \frac{1.5}{4}$), such that:

$$N_i^{t+1} = N_i^t - c_i^t + \frac{1.5}{4} \sum_{j=1}^{4} c_j^t$$

Instead of receiving fresh endowments every period, participants receive one initial endowment only at the beginning of the first period. A participant's subsequent endowment equals her profit from the current period such that a decision in one period has consequences on her future endowments. For this reason, the game is described as a *dynamic* public good game

2.3. Sample

I recruited the participants from the so called *HamburgPanel* using HROOT (Bock et al., 2014). The panel is provided by the University of Hamburg's Research Laboratory, which used a randomized last digits approach to build the panel while drawing from the population of citizens from Hamburg, Germany.

Describe sample properties here.

2.4. Software

The experiment was created using oTree (Chen et al., 2016) and can be found on GitHub.

3. Results

3.1. Online Feasibility

How did the participants, who have never participated in a online group experiment before, cope with the situation?

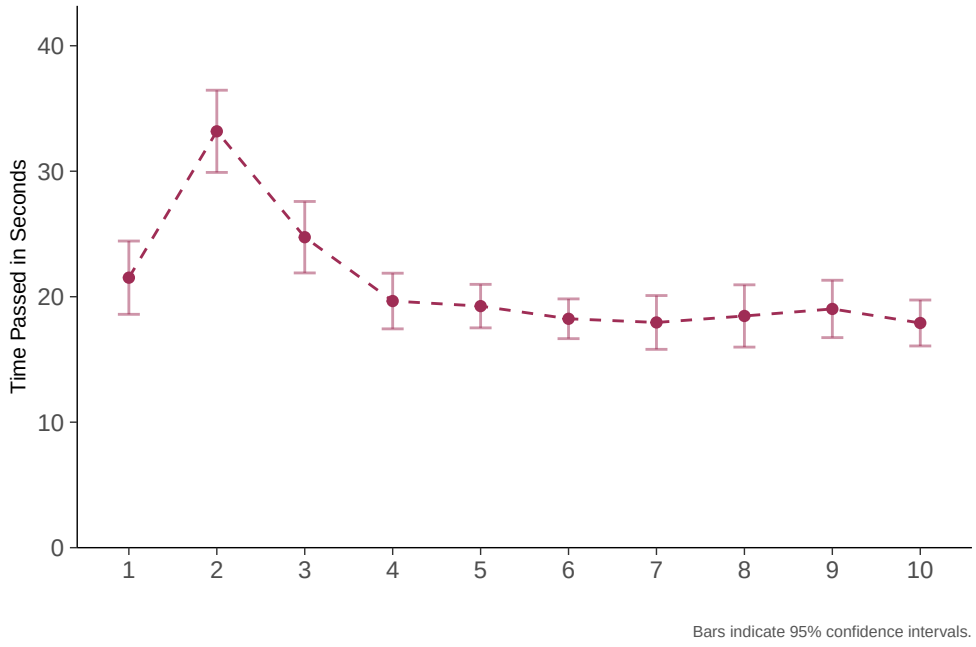

Figure 1: Average Time Spent for each Contribution per Period

3.2. Pre-registered GMTV Replication

Result 1. *The* `NOPUNISH 10` *treatment of Gächter et al. (2017) can be replicated because the replication data resemble the original data with respect to initial and final contributions, wealth and growth as well as inequality between and within groups.*

This is remarkable given the different sample and language, the different software and user interface as well as the online setting during the COVID19 pandemic.

3.2.1. Contributions

First, I ask whether the samples differ with respect to their initial contributions to the public good. Is our replication sample more pro-social than the original sample? Figure 1 reveals that it is not. The distributions of both samples look fairly similar. Both samples contributed 10 tokens, that is, 50% of their endowments on average (median and mean).[1] Moreover, both samples' initial contributions resemble initial contributions participants usually make in the standard game with partner matching.[2] However, in the dynamic game presented here, we are particularly interested in the subsequent periods because differences add up exponentially. Do the two groups remain similar over the course of time?

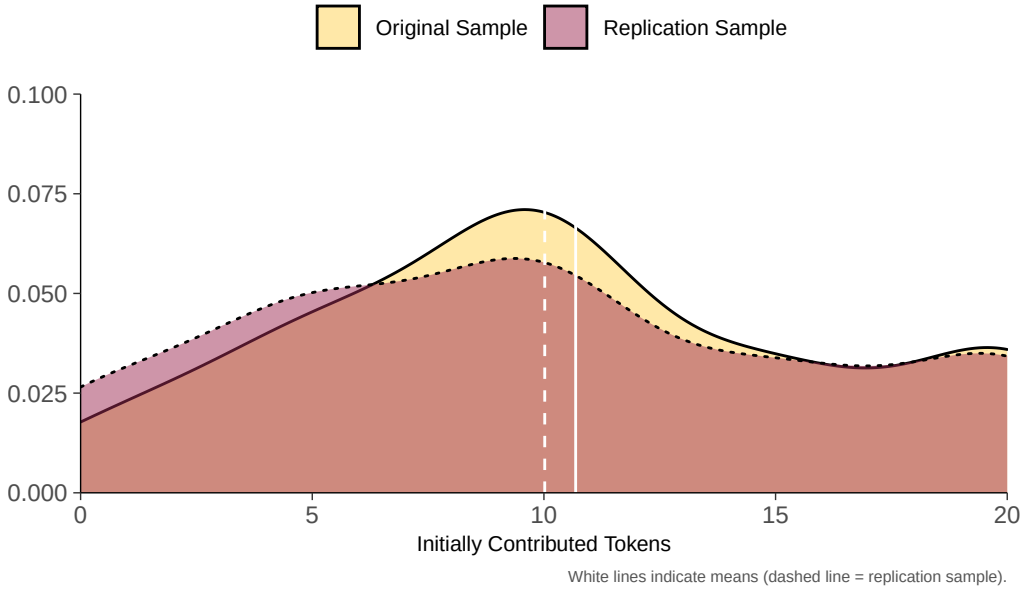

Figure 2: Individual contributions to the dynamic public good in the first period

In particular, do the two samples' contributions follow the same path over the 10 periods they played? The answer is *no*. Figure 2 illustrates that the samples make similar contributions at the beginning and the end of the game but behave differently in between. More precisely, the left panel–depicting the average contributions in absolute terms–shows that the original sample contributed substantially more than the replication sample *in all but the first and last periods*. For this reason, the original sample's behavior differs from the replication sample's behavior in two aspects: it contributes more and exhibits a considerable drop in the last period (whereas the replication sample's contributions flatten).

Note that increasing contributions over time imply increasing endowments over time. Hence, absolute contributions do not us much about the willingness to cooperate. For this reason, the right panel in Figure 2

[1] The two-sided rank sum test (comparing differences between samples) yields a p-Value of 0.3926 for the mean contribution in first round of the game.

[2] See Figure 3B in Fehr and Gächter (2000) (p.989), for instance.

shows the average *share of endowments contributed* over time. Both samples exhibit a similar pattern: they decline and do not stabilize. However, both samples also differ with respect to one aspect: the replication sample's share of contributions declines faster.

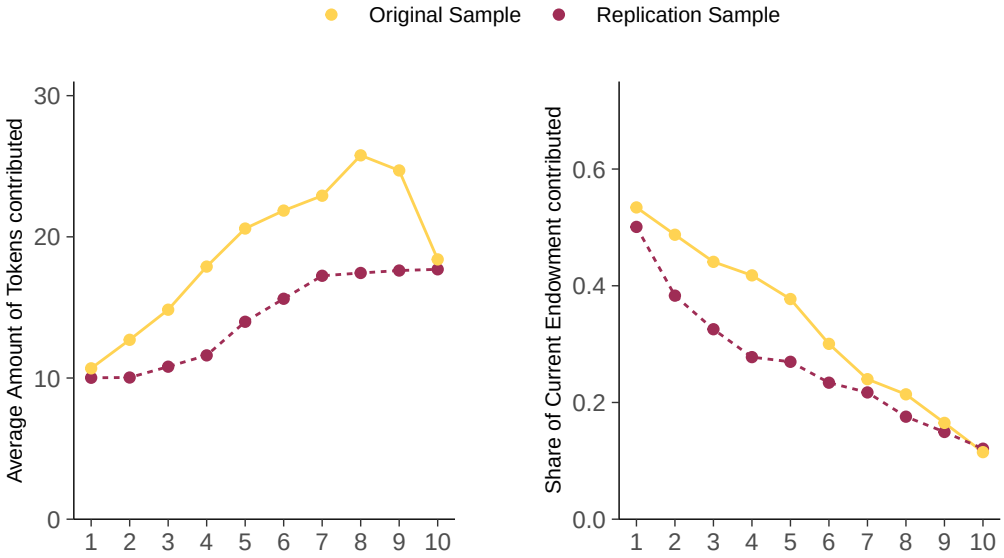

Figure 3: The average amount of tokens contributed over time in treatments.

Again, both samples' behavior resembles the contributions participants usually make in the standard game with partner matching: contributions equal approximately half of endowments in the very first period and decrease to around 10% of endowments by the last period.[3] In the dynamic game presented here, however, different paths lead to different levels of wealth – even if they share the same start- and end-points. I am thus, more interested in the contributions' implications for wealth generation and growth.

3.2.2. Wealth Creation

How do the different contribution-paths translate into wealth?[4] Given that the original sample contributed more in most of the periods, one would expect the respective groups to be considerably more wealthy. Figure 3 indicates just that. The grey lines show that an average group in the original sample accumulated about 478 tokens. In contrast, an average group in the replication sample accumulated about 380 tokens. This difference is insignificant at conventional levels[5] though.

Although there clearly is growth, groups do not realize the maximal potential efficiency: under full cooperation, a group can accumulate at least 4613 tokens or EUR 230. This is depicted in the left panel of Figure 4, where one can see the average wealth over time by sample. The panel illustrates for both samples that growth was continuous and surprisingly linear, given the exponential character of the game's design. To sum up, the contribution behavior differed between samples. In contrast, neither the eventual wealth nor the corresponding growth paths differed. Differences in contribution behavior did, thus, not translate in significantly different wealth outcomes.

Why? Perhaps because the heterogeneity within samples and across groups has been too large to *detect* a significant difference. The right panel of Figure 4 depicts heterogeneity: In the replication sample, the

[3]The right panel is thus, comparable to the visualizations *and results* in the standard game. See, for instance, Figure 1B in Fehr and Gächter (2000) (p.986).

[4]To measure wealth and growth, I define a variable called *stock* which sums the endowments of all participants in a given group at the end of the round (that is, after the contributions have been made, multiplied and redistributed).

[5]The two-sided rank sum test (comparing differences between samples) yields a p-Value of 0.1356 for the mean stock in last round of the game.

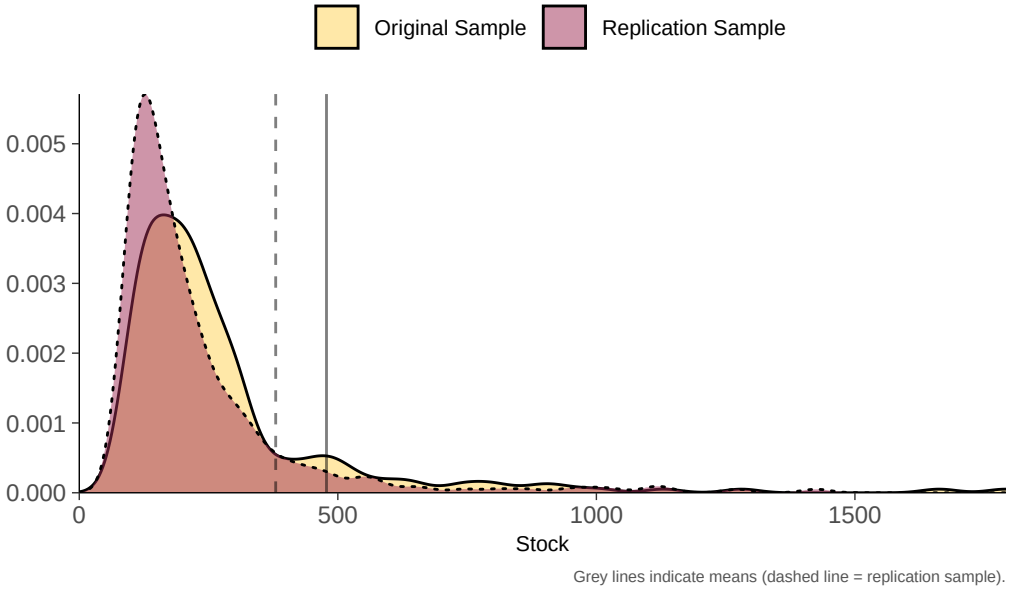

Figure 4: Groups' income at the end of the game

richest group earned 1425 tokens (which is about 1781% of the initial endowment) whereas the poorest group ends up with 92 tokens (115%). More broadly, the replication sample is characterized by inequality between groups ($SD_{Replication} = 336.06$). The same holds true for the original sample ($SD_{Original} = 393.58$). Hence, the heterogeneity across groups does not differ between samples, which is remarkable because the replication sample was drawn from a more heterogeneous (non-convenience sample). Does it differ within groups?

3.2.3. Inequality

Given the different samples and the possibility of endogenous growth–which essentially is the main feature of the game–I ask whether and how the inequality grows *within* groups. Figure 5 illustrates that inequality did grow: at the end of the game, the original and the replication groups exhibit an average Gini coefficient of 0.23 and 0.22, respectively.[6] Because every participant started with the same initial endowment (in *Period 0*, so to speak), every group started equally–with a Gini coefficient equaling zero.

Figure 6 shows that this initial state of equality ended with the first period already: both samples exhibit a stark incline in inequality before the second period started. From then on, the respective Gini coefficients grew slowly but continuously – for both samples.

3.3. Generalizability

4. Conclusion

5. Appendix

5.1. A: Pure Replication

5.2. B: Growth and *Inequality (exploratory)*

In contrast to Gächter et al. (2017), I did not ask how rich and poor groups differ. Instead, I was wondering, whether equal groups are wealthier. More precisely, does the Gini coefficient correlate with growth and wealth creation? To answer that question, Figure 7 applies a median split showing equal and unequal groups' wealth over time.

[6]The two-sided rank sum test (comparing differences between samples) yields a p-Value of 0.6059 for the mean Gini coefficient in last round of the game.

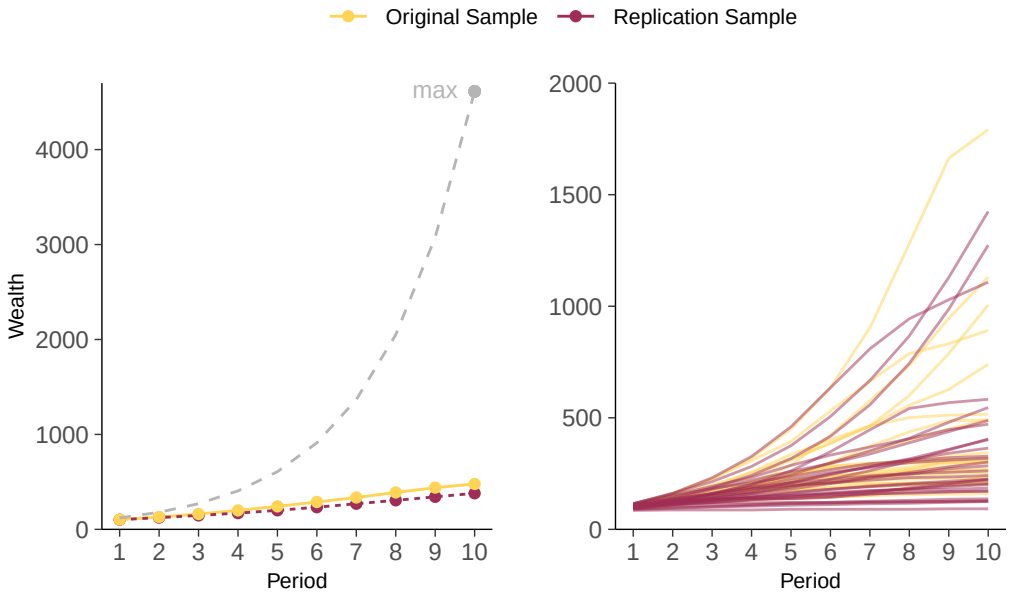

Figure 5: Average wealth over time across samples.

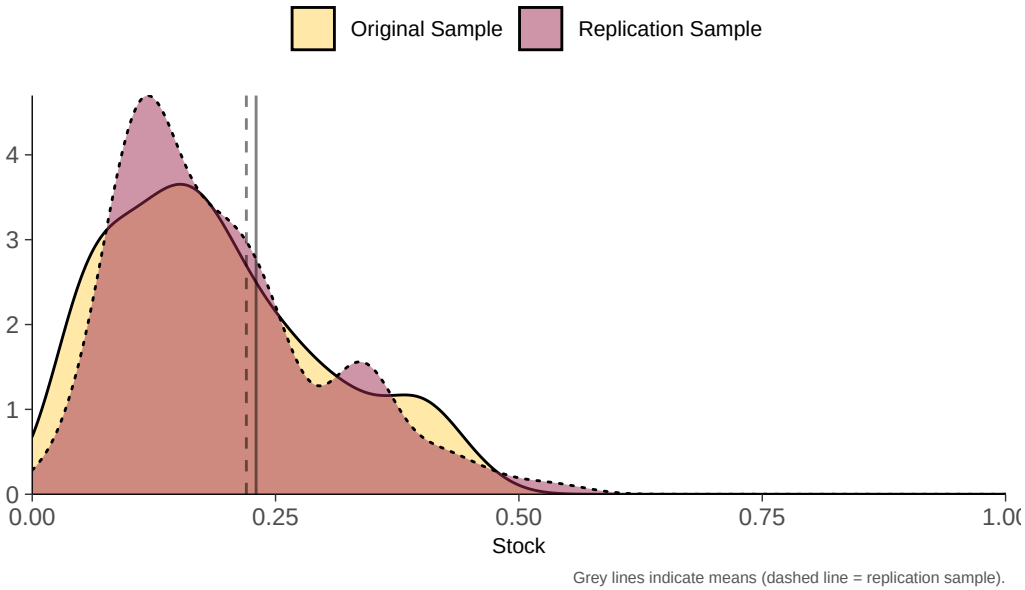

Figure 6: Groups' Gini coefficients (within groups) at the end of the game

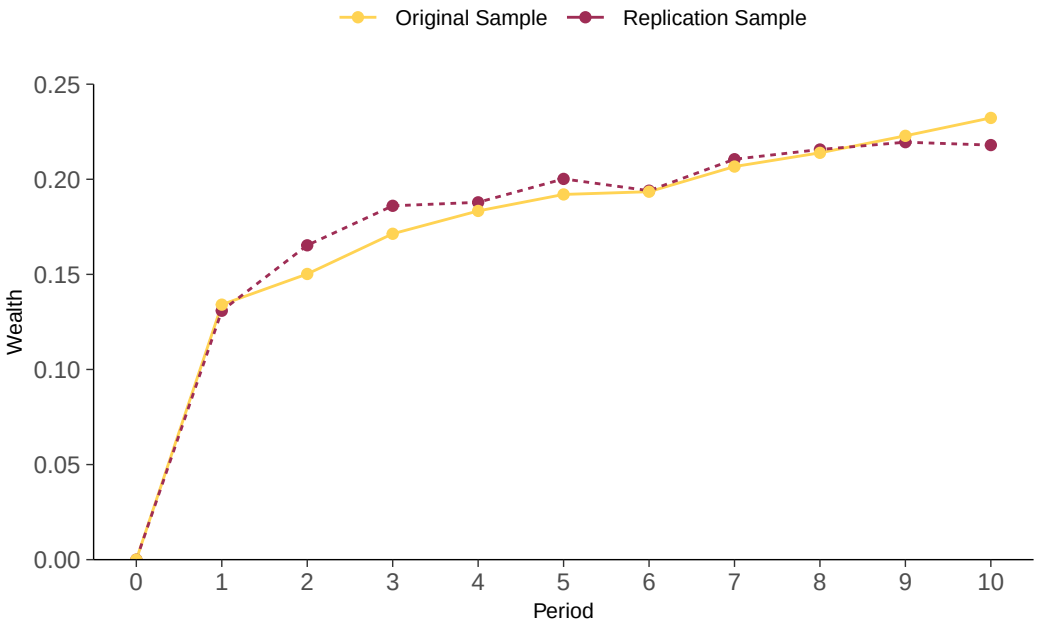

Figure 7: Average Gini coefficient (within groups) over time across samples

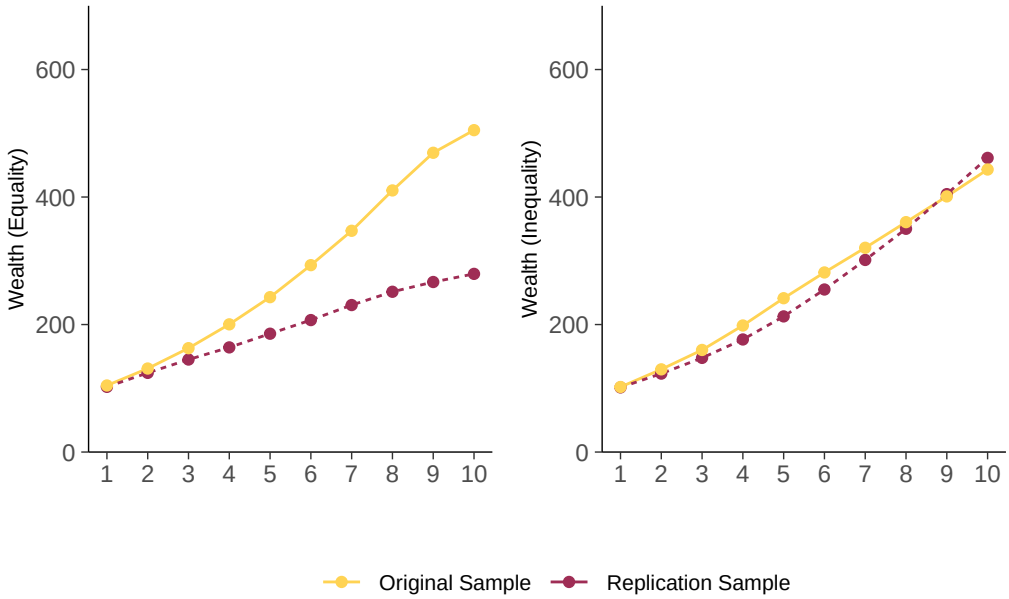

Figure 8: Average wealth over time across treatments.

References

Bem, D.J., 1987. Writing the empirical journal article, in: The compleat academic. Psychology Press, pp. 171–201. URL: https://www.taylorfrancis.com/chapters/edit/10.4324/9781315808314-10/writing-empirical-journal-article-daryl-bem, doi:https://doi.org/10.4324/9781315808314-10.

Bock, O., Baetge, I., Nicklisch, A., 2014. hroot: Hamburg registration and organization online tool. European Economic Review 71, 117–120. URL: https://www.sciencedirect.com/science/article/pii/S0014292114001159, doi:https://doi.org/10.1016/j.euroecorev.2014.07.003.

Chen, D.L., Schonger, M., Wickens, C., 2016. otree-an open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance 9, 88 – 97. URL: http://www.sciencedirect.com/science/article/pii/S2214635016000101, doi:10.1016/j.jbef.2015.12.001.

Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. American Economic Review 90, 980–994. URL: https://www.aeaweb.org/articles?id=10.1257/aer.90.4.980, doi:10.1257/aer.90.4.980.

Gächter, S., Mengel, F., Tsakas, E., Vostroknutov, A., 2017. Growth and inequality in public good provision. Journal of Public Economics 150, 1–13. URL: https://www.sciencedirect.com/science/article/pii/S0047272717300361, doi:10.1016/j.jpubeco.2017.03.002.

Hamermesh, D.S., 2007. Viewpoint: Replication in economics. Canadian Journal of Economics/Revue canadienne d'économique 40, 715–733. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2966.2007.00428.x, doi:https://doi.org/10.1111/j.1365-2966.2007.00428.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2966.2007.00428.x.

Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. American Economic Review 92, 1644–1655. URL: https://www.aeaweb.org/articles?id=10.1257/000282802762024700, doi:10.1257/000282802762024700.