# Heterogeneity in Public Good Contributions: The Pivotal Role of the First Round

**Mehmet Y. Gürdal**[*] *Boğaziçi University* **Orhan Torul**[†] *Boğaziçi University* **Mustafa Yahşi**[‡] *Koç University*

June 2020

**Abstract**

In this paper, we study heterogeneity in public good contributions across groups in Istanbul. We experimentally demonstrate that groups that start at relatively higher levels of cooperation in the initial round of the game succeed in reaching high contribution rates by the end of the game, as well. On the other hand, groups that start at low contribution levels are unable to generate a substantial increase in contributions over periods. We show that an agent-based model employing the first-period contributions and simple decision rules based on the experimental data is quite successful in capturing the evolution of contributions in distinct groups in Turkey. Next, we show using data from Herrmann et al. (2008) that Istanbul does not single out regarding its heterogeneity in first-period average contributions and their persistence throughout the game. Instead, these stylized patterns are common across cities that share socio-economic similarities with Istanbul. We conclude that concentrating only on city-level averages and ignoring within-city heterogeneity across groups leads to misleading inferences for several places around the globe.

**Keywords:** Public goods experiment; Punishment; Cooperation; Culture
**JEL Classification:** C91; C92; D81; J16

[*]Corresponding Author. Address: Boğaziçi University, Department of Economics, 34342 Bebek, Istanbul, Turkey. E-Mail: mehmet.gurdal@boun.edu.tr

[†]Address: Boğaziçi University, Department of Economics, 34342 Bebek, Istanbul, Turkey.
E-Mail: orhan.torul@boun.edu.tr

[‡]Address: Koç University, Department of Economics, Rumelifeneri Yolu 34450 Sarıyer, Istanbul, Turkey.
E-Mail: myahsi15@ku.edu.tr

# 1 Introduction

In their most classical form, public good experiments involve a group of subjects being matched in small groups and contributing towards a non-excludable group account over multiple periods. When the group size is $n$ and one unit contribution to the group account increases each member's earnings by $p$, with $p < 1$ and $np > 1$, the Nash equilibrium is zero contribution by all members, whereas the efficient outcome is the full contribution. In these experiments, subjects generally start with non-zero contributions and converge to the equilibrium of zero contributions as the game evolves (Isaac and Walker, 1988).

In their seminal studies, Fehr and Gächter (2000, 2002) implement a design where subjects play a game of known length, observe each others' contributions each period and decide how much to deduct from the earnings of other group members, i.e. *sanction* each other.[1] This deduction is costly for the subject, therefore it is classified as a form of strong reciprocity (*costly altruism*). When subjects are allowed to assign punishment points to their group members, this can act as a facilitator of cooperation and keep the contribution levels high in the experiment.

An extant body of literature aims to explore location-specific patterns in cooperation with and without sanctioning. Among these, Herrmann et al. (2008) employ the design by Fehr and Gächter (2000) and compare the success of the punishment mechanism among 1120 subjects in 16 different cities around the world. They demonstrate that the success of subject groups in using the sanctioning mechanism as an effective tool to sustain cooperation varies greatly across different subject pools. Also, certain subject groups engage in a substantial amount of antisocial punishment, which occurs when a subject retaliates against a group member who contributed more than her to the group account. The extent of antisocial punishment is negatively correlated with contribution levels, and subject pools where antisocial punishment is rare successfully manage to reach high contribution rates.

Istanbul is one of the cities that Herrmann et al. (2008) include in their subject pool. It is characterized by relatively higher levels of antisocial punishment and relatively low success in using the punishment mechanism as a facilitator of cooperation. Indeed, Herrmann et al. (2008) do not find a significant difference in contributions in games with or without punishment opportunity, as in the case of cities that exhibit similarities with Istanbul. In this paper, we report the results of an experiment based on the design by Fehr and Gächter (2000). While the experiment is conducted in the *same school* as in Herrmann et al. (2008), we employ a *between-subject* design, whereas they

[1] Earlier examples where a sanctioning mechanism is used in similar games involve Yamagishi (1986) where subjects could punish the least cooperative group member and Ostrom et al. (1992) where there is a fixed cost of sanctioning and being sanctioned. Since the game length is unknown by the subjects in the latter study, it is not possible to rule out the incentives for reputation behind contribution and sanctioning decisions.

employ a *within-subject* design. We find, contrary to what Herrmann et al. (2008) find, that the punishment mechanism significantly increases contributions and a substantial fraction of groups manage to reach contribution rates of 80% or more of the endowment. However, the overall average contribution rates (around 55%) under punishment still fall short of those observed among subjects pools from cities like Boston, Copenhagen, St. Gallen, Zurich, and Nottingham.[2] Initial contributions under punishment exhibit a great degree of variation across groups and correlate positively with cooperation in further periods. While antisocial punishment is quite rare, it has a detrimental effect on future contributions.

Our main contribution is the demonstration of *extensive heterogeneity across groups* where punishment is available. In particular, we show that groups that start at relatively higher levels of cooperation in the initial round of the game manage to keep average contribution relatively high and most groups succeed in even raising it, with some groups reaching full contribution levels at the end of the game. On the other hand, groups that start at low levels are unable to generate a substantial increase in the contributions. Burlando and Guala (2005) underline the case of heterogeneous agents scenario as the reason behind the variation in contributions across groups and state that their results are in line with Andreoni and Miller (2002). Accordingly, we claim that the first-period contribution is the most fundamental element to drive heterogeneity in late contribution rates across groups. We show that an agent-based model, employing the first-period contributions and simple decision rules based on the experimental data is quite successful in capturing the evolution of contributions in distinct groups in Turkey.[3]

Next, we show using data from Herrmann et al. (2008)'s P-experiment that Istanbul does not single out with its idiosyncrasy vis-à-vis heterogeneity in first-period average contributions and their persistence throughout the game. Instead, these stylized patterns are common across cities that share socio-economic similarities with Istanbul, particularly in Athens (Greece), Dnipropetrovsk (Ukraine), Samara (Russia), Minsk (Belarus), Riyadh (Saudi Arabia), and Muscat (Oman). Accordingly, we conclude that concentrating only on city-level averages and ignoring within-city heterogeneity across groups result in misleading public good contribution inferences for several places around the globe.

The rest of the paper is organized as follows: in section 2 we discuss the details of the experiment, in section 3 we report our empirical findings, in section 4 we describe our agent-based model

[2] In all of these cities, the average contribution rate is 75% or more of the endowment when punishment is available.

[3] Arifovic and Ledyard (2012) use elements from both individual evolutionary learning (IEL) and other-regarding preferences (ORP) to propose a sophisticated theoretical model (IELORP) that can deliver the stylized facts observed in voluntary contribution mechanisms. Lee-Panagos (2016) extends the already-sophisticated model by Arifovic and Ledyard (2012) so as to account for developments in the stylized facts of voluntary contribution mechanisms. We, instead, propose a very parsimonious agent-based model that concentrates on the primary focus of this paper: how the first-period contributions shape the rest of the rounds.

and present our resultant findings, in section 5 we compare our findings in detail with those by Herrmann et al. (2008), and section 6 concludes.

## 2 Experiment

The experiment is based on the design by Fehr and Gächter (2000) and involves two treatments. The treatment termed as the N-experiment involves subjects being matched in groups of four and interacting within the same group throughout 10 periods. Each period, subjects are endowed with an initial endowment of 20 tokens, out which they can contribute to a "group project". For every token invested in the group project, each group member earns 0.4 tokens. Subjects' period earnings are the sum of earnings from the group project and the part of the endowment that is not invested in the group project. The treatment termed as the P-experiment is built on the N-experiment, but besides the contribution stage, it involves an additional stage where subjects see the contributions (but not the identities) of all other members in their group and can assign punishment (sanction) points. Each punishment point costs 1 token to the subject and reduces the target's earnings by 3 tokens. The total reduction on a subject's earnings is limited to the earnings from the contribution stage.

The experiment is conducted with 120 subjects in a total of 10 sessions at the economics laboratory of Boğaziçi University. Each session involves 12 subjects. We implemented a between-subject design where 60 subjects only take part in N-experiment and the other 60 subjects only take part in the P-experiment. An e-mail was sent to subjects who previously indicated an interest in joining economics experiments, and subjects could register online for a date and time of their choosing. No subject participated more than once and the sessions lasted 45 minutes on average. Subjects were paid in cash at the end of the experiment. The exchange rate is 0.1 Turkish liras per token in the current study.[4]

## 3 Results

In Table 1, we provide summary statistics for the observed contribution levels. The contributions in the current experiment start at very close levels (around 9) in both the P-experiment and the N-experiment. However, as of the $10^{\text{th}}$ period, the contributions diminish to very low levels (2.85) in the N-experiment while they moderately rise to a mean value of approximately 12 in the P-experiment. We find a significant difference between the P-experiment and the N-experiment when

[4]At the time of the experiments (November 9-11, 2015), 1 Turkish lira corresponded to approximately \$0.35 United States dollars.

all periods are considered ($p = 0.019$), but not for Period 1 ($p = 0.950$). Considering all periods, the computation of statistical power analysis on the respective non-parametric test to reject the equality of the average contributions in the N-experiment and the P-experiment yields a power value of $pw = 0.78$.[5]

Herrmann et al. (2008) also employ the design by Fehr and Gächter (2000) to measure the performance of the punishment mechanism among 1120 subjects in 16 different cities around the world, including Istanbul. However, while we implement a *between-subject* design, they use a *within-subject* design where the P-experiment follows the N-experiment for a majority of subject pools in their sample, Istanbul included. We report the contributions of subjects in Istanbul, Boston, and Copenhagen from that study in Table 1.[6]

Using the data from Istanbul, Herrmann et al. (2008) find a difference between the two treatments in Period 1 (with the P-experiment generating lower contributions than the N-experiment), but no significant difference when all periods are considered. The summary statistics and non-parametric test results suggest that a between-subject design increases the effectiveness of the punishment mechanism in Istanbul, yet, compared to the results from Boston and Copenhagen, subjects in the current study still contribute at lower rates on average even under the existence of punishment. In addition, compared to Boston and Copenhagen, the availability of the punishment mechanism results in much lower average earnings in Istanbul, both for the current study and for Herrmann et al. (2008).[7]

**Table 1:** Mean Contributions and Earnings

| | Contribution in Period 1 | | | Contribution in All Periods | | | Earnings | |
|---|---|---|---|---|---|---|---|---|
| | N | P | $p$-value | N | P | $p$-value | N | P |
| Istanbul - Current study | 8.9 | 9.1 | 0.950 | 6.1 | 10.9 | 0.019 | 23.6 | 19.4 |
| Istanbul - Herrmann et al. (2008) | 8.9 | 6.5 | 0.034 | 5.4 | 7.1 | 0.326 | 23.3 | 17.0 |
| Boston - Herrmann et al. (2008) | 13.0 | 16.0 | 0.012 | 9.3 | 18.0 | 0.002 | 25.6 | 27.9 |
| Copenhagen - Herrmann et al. (2008) | 14.1 | 15.4 | 0.088 | 11.5 | 17.7 | 0.001 | 26.9 | 27.7 |

*Notes:* The table reports the average contributions (out of 20 tokens) and the p-values for Mann-Whitney tests that use group average contributions as independent observations and test for the equality of contributions in the N-experiment and the P-experiment.

[5] Since the power value of the test is reasonably high at 0.78, it is safe to assume that there is a significant difference in the average contributions between the P-experiment and the N-experiment with our sample sizes. For a detailed discussion on statistical power analysis, see Cohen (1988).

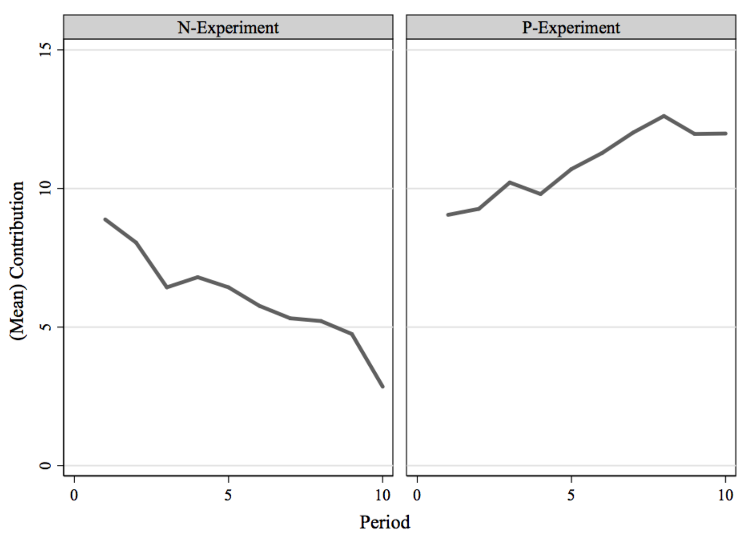
[6] The latter two cities are the ones where the highest average contributions in the P-experiment or the N-experiment were reported, respectively.

[7] For a detailed comparison, see section 5.

### 3.1 N-Experiment

The overall behavior in the N-experiment exhibits the classical pattern of decaying contributions. In Figure 1, we present the evolution of contributions averaged over groups for the N-experiment and the P-experiment. During the first half of the N-experiment, the frequency of subjects who contribute zero is 27%, whereas in the second half this frequency reaches 46%. Among the 15 distinct groups in this experiment, 11 groups end up at very low (3 tokens or fewer) average contribution levels, 3 groups end up at moderate (between 6.75 and 8.5 tokens) contribution levels, and only a single group manages to reach a contribution level of 12 by the $10^{th}$ period. These values point out to some degree of heterogeneity, but clearly, the most dominant pattern observed in the N-Experiment is the decline of contributions over periods. On the contrary, the *evolution* of contributions in the P-Experiment exhibits a great deal of heterogeneity among groups, as we discuss in detail in the next section.

**Figure 1:** Timeline of Average Contribution by Treatment




*Notes:* The left panel reflects the average contribution in N-experiment while the right panel shows the respective statistic in P-experiment.

### 3.2 P-Experiment

The way group members update their contributions in the P-experiment follows a close relationship with the respective group average. In particular, when a group member observes that her contribution falls short of the group average, she will seldom decrease her contribution next period. Instead, she will almost always either increase her contribution or keep it the same. Similarly, when her

**Table 2:** Contribution Change by Group Average

| Group Average | Next Period Contribution | | |
|---|---|---|---|
| | Increases | Remains the Same | Decreases |
| Lower | 39 | 81 | 116 |
| Equal | 9 | 53 | 5 |
| Higher | 161 | 45 | 31 |

*Notes:* The table reports the frequency of next-period contribution changes by subjects when they compare their current-period contribution to the group average.

contribution exceeds the group average, the next period contribution will very frequently be either lower or the same. In table 2, we report the frequencies of the direction of change in contributions conditional on how one's last period contribution compares to the group average.

#### 3.2.1 Decision to Change Contributions

In a series of logit regressions reported in Table 3, we first set out to understand the dynamics behind the decision to revise (increase or decrease) contributions in the P-experiment. Here, we also control for certain attitudinal and demographic information about subjects and capture group-level fixed effects.[8]

In Model 1 and Model 2, the dependent variable, *change*, takes the value of 1 if a subject's contribution in the current period differs from her contribution in the previous period, and 0 otherwise. Our estimations reveal that the only variable that has a significant effect on the *change* variable is the punishment points received from other group members in the previous period (Model 1) and this effect is robust to controlling for other factors (Model 2). In Model 3 and Model 4, the dependent variable, *increase*, takes the value of 1 if a subject's contribution in the current period is larger than her contribution in the previous period, and 0 otherwise. In these models, we restrict our working sample to a subset of observations featuring only subjects who contribute less than the group average in the previous period. Once again, the only variable with a significant effect is the number of punishment points received from other group members in the previous period (Model 3), and this effect is robust to controlling for other factors (Model 4). In Model 5 and Model 6, the dependent variable, *decrease*, takes the value of 1 if a subject's contribution in the current period is less than her

[8]The set of control variables used in regressions involve the following: age in years, a gender dummy, number of older siblings, number of younger siblings, a dummy for being an only child, binary response to the trust question ("Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?"), Likert response to the risk question ("How willing are you to take risks in general?"), number of economic classes taken (censored at 4), number of friends in the session, an indicator for membership in an organization, Likert response to the question "How reliable do you think your responses are in this experiment?"

contribution in the previous period, and 0 otherwise. In these models, we again restrict our working sample to a subset of observations, this time featuring only subjects who contribute more than or equal to the group average in the previous period. In these estimations, the average contribution of other group members in the previous period and contribution in the previous period have a significant effect on the next period contributions (Model 5). When controlling for other factors (Model 6), punishment points received from other group members in the previous period yields a significant effect, as well.

#### 3.2.2 Magnitude of Change in Contributions

Next, we run a series of ordinary least squares (OLS) regressions and report our findings in Table 4. Here, we use *contribution* as the dependent variable and restrict our working sample to observations featuring subjects who revise their contribution in the current period. Models 1 and 2 involve all subjects who either raised or reduced their contributions (but not those who kept it the same). The number of punishment points received from other group members in the previous period, the average contribution of other group members in the previous period, and the last period contribution have a significant effect on the next period contributions (Model 1). The effect of these variables preserves significance when controlling for other factors (Model 2). In Models 3 and 4, we restrict our working sample to observations featuring subjects who contribute less than the group average in the previous period and raise their contributions in the current period. Here, the number of punishment points received from other group members in the previous period, the average contribution of other group members in the previous period, the last period contribution, and the period variable have a significant effect on the next period contributions (Model 3). The significance of these variables (except period) is robust to additional controls (Model 4). In Models 5 and Model 6, we restrict our working sample to observations featuring subjects who contribute more than or equal to the group average in the previous period and reduce their contributions in the current period. The number of punishment points received from other group members in the previous period, the average contribution of other group members in the previous period and the last period contribution have a significant effect on the next period contributions (Model 5). Not surprisingly, received punishment points have a negative effect on contributions since they constitute antisocial punishment in this setting. The effect of these variables preserves significance when controlling for other factors (Model 6).

**Table 3:** Determinants of Contributions: Logit

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Received Points in $t-1$ | 0.314** | 0.249** | 0.175* | 0.214** | 0.111 | 0.190** |
| | (0.124) | (0.108) | (0.100) | (0.093) | (0.076) | (0.089) |
| Other's average contribution in $t-1$ | -0.047 | -0.047 | 0.043 | 0.057 | -0.318*** | -0.342*** |
| | (0.044) | (0.038) | (0.064) | (0.047) | (0.066) | (0.059) |
| Contribution in $t-1$ | 0.017 | -0.005 | 0.035 | -0.010 | 0.216*** | 0.265*** |
| | (0.029) | (0.037) | (0.077) | (0.060) | (0.040) | (0.046) |
| Period | 0.084 | -0.020 | -0.068 | -0.006 | -0.018 | -0.039 |
| | (0.072) | (0.055) | (0.054) | (0.075) | (0.067) | (0.078) |
| Final period | -0.395 | -0.067 | -0.034 | -0.227 | 0.376 | 0.392 |
| | (0.288) | (0.242) | (0.573) | (0.527) | (0.351) | (0.468) |
| Constant | No | No | No | No | No | No |
| Controls | No | Yes | No | Yes | No | Yes |
| $N$ | 540 | 540 | 237 | 237 | 303 | 303 |
| $adj.R^2$ | | | | | | |

*Notes:* Standard errors are clustered across different groups, reported in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

**Table 4:** Determinants of Contributions: OLS

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Received Points in $t-1$ | 0.244** | 0.191* | 0.399*** | 0.324** | -0.513** | -0.363** |
| | (0.089) | (0.108) | (0.107) | (0.119) | (0.187) | (0.153) |
| Other's average contribution in $t-1$ | 0.612*** | 0.608*** | 0.473*** | 0.426*** | 0.495*** | 0.565*** |
| | (0.067) | (0.067) | (0.090) | (0.107) | (0.159) | (0.162) |
| Contribution in $t-1$ | 0.336*** | 0.256*** | 0.511*** | 0.493*** | 0.412*** | 0.341** |
| | (0.073) | (0.073) | (0.097) | (0.099) | (0.118) | (0.125) |
| Period | 0.039 | -0.031 | 0.151* | -0.028 | -0.107 | -0.111 |
| | (0.061) | (0.125) | (0.082) | (0.124) | (0.069) | (0.207) |
| Final Period | -0.527 | -0.262 | -0.649 | 0.007 | 0.471 | 0.198 |
| | (0.603) | (0.768) | (0.528) | (0.633) | (1.228) | (1.601) |
| Constant | No | No | No | No | No | No |
| Controls | No | Yes | No | Yes | No | Yes |
| $N$ | 361 | 361 | 161 | 161 | 121 | 121 |
| $adj.R^2$ | 0.895 | 0.901 | 0.956 | 0.961 | 0.880 | 0.898 |

*Notes:* Standard errors are clustered across different groups, reported in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

**Table 5:** Social versus Antisocial Punishment

| Sender's Contribution | Punishment | | | |
|---|---|---|---|---|
| | Non-zero | Zero | Mean | Median |
| Higher | 338 (52.3%) | 308 (47.7%) | 1.09 | 1 |
| Not Higher | 204 (17.7%) | 950 (82.3%) | 0.33 | 0 |

*Notes:* The table reports the frequency of social and antisocial punishments along with the mean and median statistics.

#### 3.2.3 Punishment Decisions

In Table 5, we report punishment frequencies along with mean punishment points conditional on relative contributions of subjects sending the punishment (sender) and target subjects (receiver). While punishment expenditures are often not high, subjects punish those who contributed less than them around 52% of the time. On the contrary, antisocial punishment is not prevalent and occurs only around 18% of the cases when a subject observes that a group member who contributed at least as much as herself. Note that the mean punishment expenditures are very close to the values obtained in Herrmann et al. (2008) for Istanbul.

Our next step involves disentangling the effects of different variables on the likelihood of engaging in punishment. Since punishment expenditures are often low and a punishment level of zero is commonly observed, we chose to use a series of probit models, as reported in Table 6, instead of linear regressions. Model 1 (Model 3) and Model 2 (Model 4) use the first-period observations in the P-experiment; and the latter involves a set of attitudinal and demographic variables as controls whereas the former does not. When we restrict our working sample to observations in which the receiver's contribution is less than that of the sender, the two strongly significant determinants of social punishment in the first period is the *contribution of the target subject* (i.e. the receiver) and the *average contribution of the two remaining group members* (Model 1). The level of significance for these variables is robust to additional controls (Model 2). Models 3 and 4 consider instances where the sender's contribution is less than or equal to the receiver's contribution in the first period. We see that only the contribution of the target subject (i.e. the receiver) has a significant effect on anti-social punishment in the first period; however, the significance of this variable vanishes when we control for other factors. Model 5 and Model 6 use all observations in the P-experiment, and the latter involves a set of attitudinal and demographic variables as controls, whereas the former does not. We document that the sender's contribution, the average contribution of the two remaining group members, the punishment points that the subjects received in the previous period, the receiver's contribution and the period variable have an effect on punishment decision (Model 5). All these variables except the sender's contribution remain significant determinants of punishment decision controlling for other factors (Model 6).

Using a linear model, Herrmann et al. (2008) find that the amount of social punishment is negatively related to the contribution of the target subject and positively related to the average contribution of remaining group members in Istanbul (Table S3A). On the other hand, they find that the amount of antisocial punishment for the same subject pool is negatively related to the punisher's contribution and the period of the experiment (Table S3B). For the current study, we obtain the same sign for all the effects, but the effect of punisher's contribution to antisocial punishment is insignificant. In addition, they don't find an effect of the punishment points received in the previous period on antisocial punishment, whereas we find a positive effect of this variable in the current study.

**Table 6:** Punishment Decisions

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Receiver's Contribution | -0.100*** | -0.151*** | -0.085*** | -0.018 | -0.146*** | -0.156*** |
| | (0.034) | (0.051) | (0.030) | (0.035) | (0.021) | (0.023) |
| Sender's Contribution | 0.007 | -0.048 | 0.026 | -0.023 | 0.033* | 0.025 |
| | (0.032) | (0.058) | (0.041) | (0.050) | (0.020) | (0.024) |
| Avg. Contr. of Remaining Members | 0.067** | 0.102** | -0.015 | 0.078 | 0.094*** | 0.106*** |
| | (0.031) | (0.042) | (0.027) | (0.049) | (0.021) | (0.020) |
| Received Sanction Points in $t-1$ | | | | | 0.045** | 0.045* |
| | | | | | (0.022) | (0.023) |
| Period | | | | | -0.082*** | -0.063** |
| | | | | | (0.028) | (0.029) |
| Final Period | | | | | 0.184 | 0.082 |
| | | | | | (0.161) | (0.188) |
| Controls | No | Yes | No | Yes | No | No |
| $N$ | 80 | 80 | 100 | 100 | 1620 | 1620 |
| $\|ln(L)\|$ | | | | | | |

*Notes:* Standard errors are clustered across different groups, reported in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

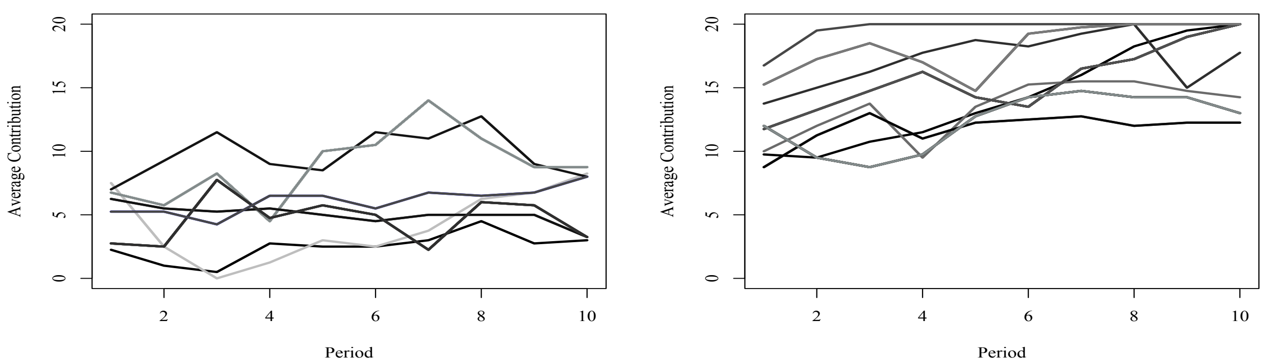
# 4 Agent-Based Modeling

In addition to differences arising from the subject design of the P-experiment, we argue that another critical factor that has first-order implications is the heterogeneity in the first round, which has long-lasting consequences.[9] Our empirical findings reveal that if subjects in Istanbul happen to start the P-experiment in a collaborative group, they sustain collaboration rates as high as in Boston and

[9]Specifically, we argue that drawing conclusions on the success rate of the P-experiment by concentrating only on *group averages* as in Herrmann et al. (2008) can be misleading, since this approach smooths out the decisive role of first-round results by aggregation.

Copenhagen. We highlight the importance of the first-period results in Figure 2, where we cluster groups with respect to their first-period average contributions (lowest 7 on the left and highest 8 on the right panel) and plot the evolution of group averages over periods. Figure 2 reveals that if a group's average contribution rate is above 43.75% in the initial round, its average contribution rate by the end of the P-experiment is no less than 61.25%.

**Figure 2:** Evolution of Average Contributions in the P-experiment




*Notes:* The figure on the left shows the average contribution over the horizon of 10 periods for 7 groups with the lowest first-period average contributions whereas the figure on the right demonstrates the statistic for the remaining 8 groups.

In order to elicit the decisive role of the first-period results, we rely on an agent-based modeling approach. In doing so, we start by taking the first-period contributions from the P-experiment data. Next, for our subjects, we impose a simple state-dependent set of decision rules motivated by the data, and we run Monte Carlo simulations for each of the 15 groups in our sample. Specifically, we proceed as follows:

1. We start our counter-factual analysis by feeding the agent-based model with *actual* first-period contribution data of our 60 subjects in the 15 groups.
2. Subjects either change their contribution or make the same amount of contribution in the next period. Following our empirical findings, we link decision rules to subjects' comparison of their last period contribution to the group average: if a subject's last period contribution is less than the group average, she either raises her contribution or keeps contributing the same amount in the next period; and if her last period contribution is at least equal to the group average, she either reduces her contribution or contributes same. In the former case, we rely on a two-layer contribution rule: in the first layer, using estimated probabilities defined over received sanction points last period $s_{i,t-1}$, we simulate if one retains her contribution or not

(Model 3, Table 3). If she does not retain, in the second layer, we impose that her contribution follows $c_{i,t} = 0.511 \times c_{i,t-1} + 0.473 \times \overline{c}_{-i,t-1} + 0.399 \times \sum_{j \neq i}^{4} s_{j,i,t-1} + 0.151 \times t$. On the contrary, if a subject's last period contribution is greater than or equal to the group average, we similarly rely on a two-layer contribution rule: in the first layer, using estimated probabilities defined over the average contribution of other members in the group during the previous period $\overline{c}_{-i,t-1}$ and last period contribution $c_{i,t-1}$, we simulate if one retains her contribution or not (Model 5, Table 3). If she does not retain, in the second layer, we impose that her contribution follows $c_{i,t} = 0.412 \times c_{i,t-1} + 0.495 \times \overline{c}_{-i,t-1} - 0.513 \times \sum_{j \neq i}^{4} s_{j,i,t-1}$.

3. We observe that subjects punish others either by a *single* sanction point or they do not punish others in 85.06% of all possible cases. We see that subject $i$'s sanctioning decision on subject $j$, $s_{i,j,t}$ depends significantly on subject $i$'s received sanction points last period $s_{i,t-1}$, the average contribution of the remaining members in the respective group in the current period $\overline{c}_{-i-j,t}$ and the round of the current period $t$, as well as her and subject $j$'s contribution in the current period: $c_{i,t}$ and $c_{j,t}$. Based on the estimated sanction probabilities from the data, we simulate if subject $i$ assigns subject $j$ a sanction point (of unity) or not (Model 5, Table 6).[10]

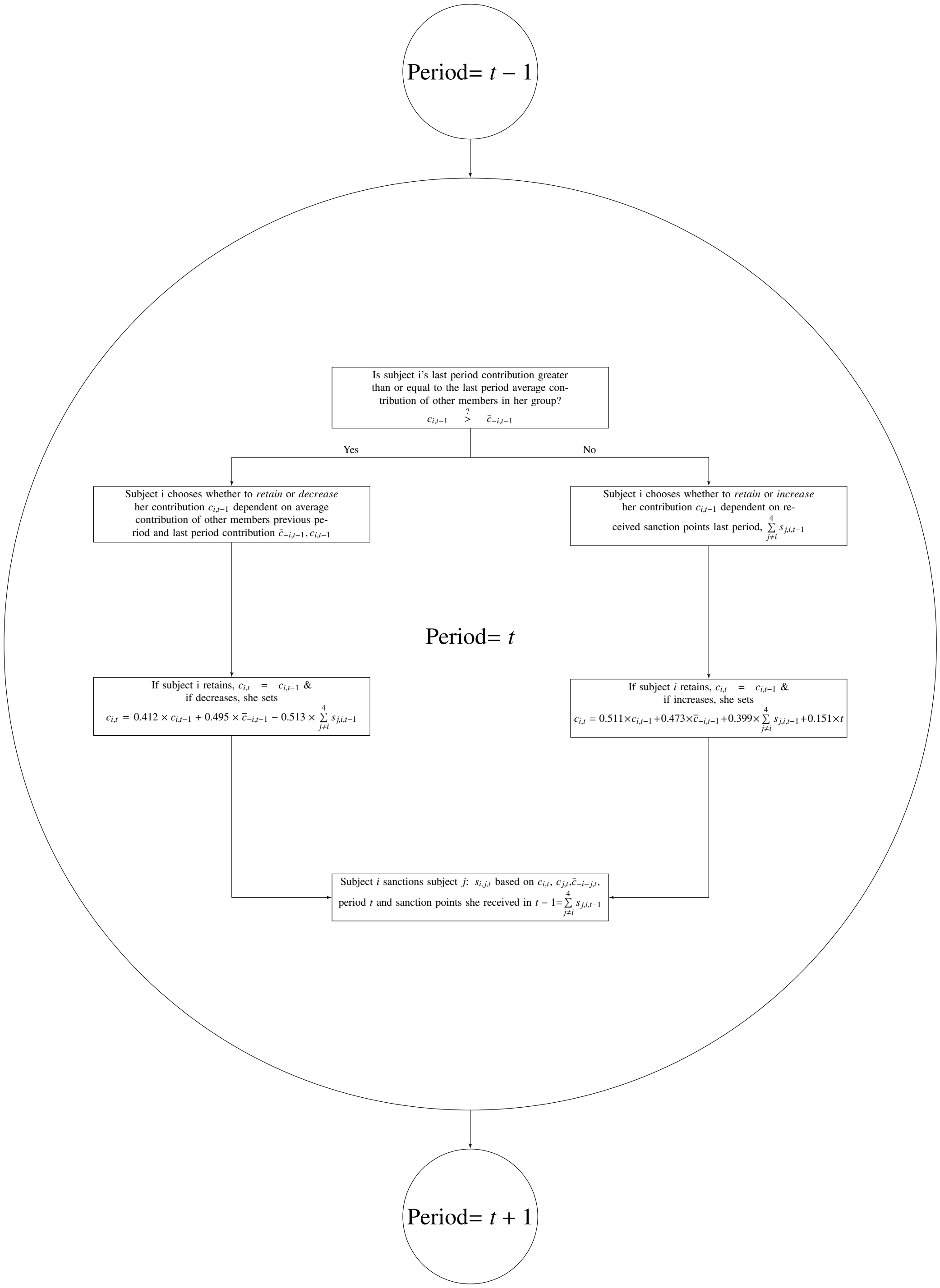
4. We move on to the next period, and recursively conduct the same steps over periods.

We summarize the timeline of the agent-based modeling algorithm in Figure 3.

We display average group contributions by our agent-based model simulations in Figure 4. In doing so, for each of the 15 groups, we plot group-specific Monte-Carlo simulation averages (dotted lines) along with their 2-standard-deviation confidence intervals (shaded grey areas), and we display them jointly with *actual* group average contributions (straight lines). Figure 4 reveals that despite its simplicity, the decision rule we impose on subjects not only mimics final round average contributions accurately but also does a fairly good job in capturing the evolution of group averages in most of the instances.

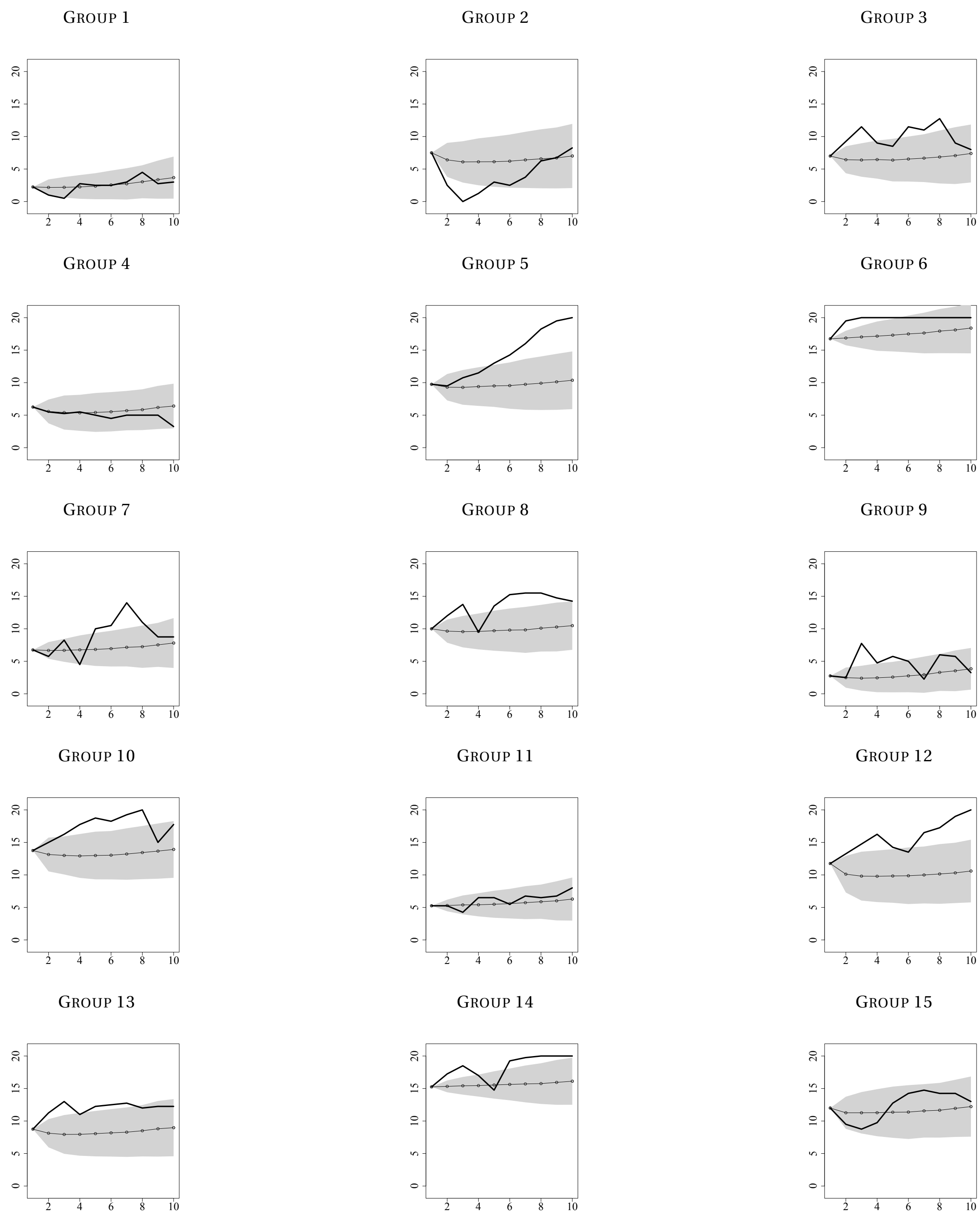
[10] We do not have a history of received sanction points in the first period; so, we proceed differently to determine the first-period sanction points. In the first-period, we see that subject $i$'s sanctioning decision on subject $j$, $s_{i,j,1}$ significantly depends on the average contribution of the remaining members in the respective group $\overline{c}_{-i-j,1}$ and $j$'s contribution $c_{j,1}$ for social punishment case (Model 1, Table 6), whereas this decision only depends on $j$'s contribution $c_{j,1}$ for anti-social punishment case (Model 3, Table 6). Based on estimated sanction probabilities from the data, we simulate if subject $i$ assigns subject $j$ a sanction point (of unity) or not.

**Figure 3:** Timeline of the Model




*Notes* Figure 3 displays the timeline of agent based modeling for the P-experiment.Only first-period contributions are retrieved from the experiment while the remaining data is constructed through the modeling. Subjects' decisions to retain or change their contribution level in subsequent periods are determined via Table 3. In the case of change, next period contributions are constructed by benefiting from Table 4. Sanctioning decisions are made via Table 6.

**Figure 4:** Agent-Based Model Simulation Results



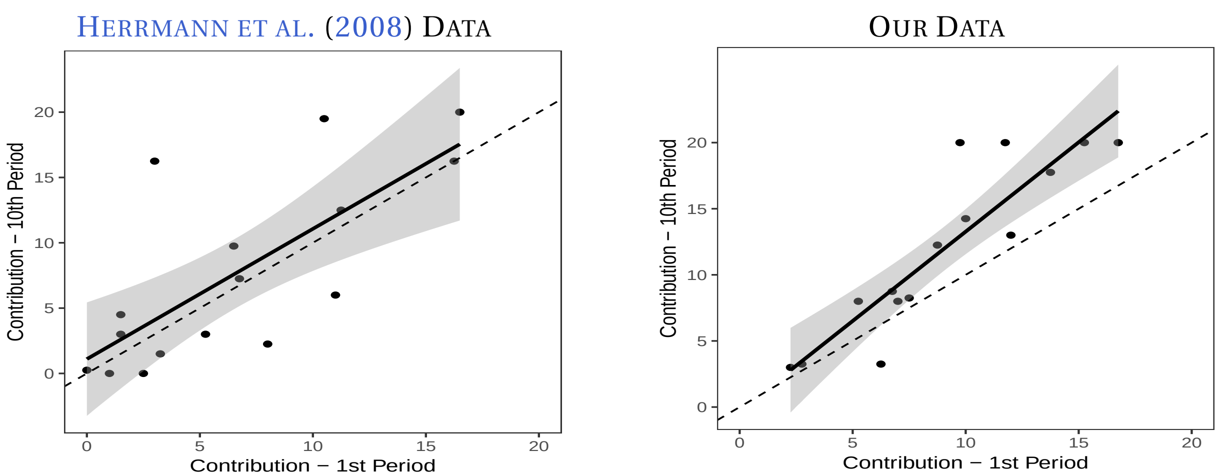

*Notes:* The horizontal axis denotes periods and the vertical axis denotes average group contribution. The solid line refers to group averages by the P-experiment data, the dotted line refers to group-specific average agent-based model results by Monte-Carlo simulations, with the shaded gray areas referring to resultant 2-standard-deviation confidence intervals.

## 5 Comparison with Herrmann et al. (2008)

Heterogeneity in first-period average contributions and its persistence throughout the game is not peculiar to our experiment. In order to demonstrate how our findings compare to those by Herrmann et al. (2008) for Istanbul, in Figure 5 we plot first-period average group contributions against their last-period counterparts by our and Herrmann et al. (2008)'s data.

Figure 5 reveals that first-period average group contributions by Herrmann et al. (2008) also exhibit a high degree of variability, as do our data. There are, however, two stark differences: First, while only 2 (out of 15) groups in our experiment contribute at a rate lower than 25% in the first period, 7 (out of 16) groups in Herrmann et al. (2008)'s experiment start the game at a contribution rate below 25%. Second, while the slope between the first and last period average group contributions via Herrmann et al. (2008)'s data is close to unity (0.997, with a standard error of 0.243), the slope via our experiment is sizeably steeper (1.352, with a standard error of 0.186). This suggests that while average group contributions by Herrmann et al. (2008) stagnated on average over time, groups in our experiment managed on average to raise their average group contributions over periods. These disparities could stem from the difference due to our *between-subject* design versus Herrmann et al. (2008)'s *within-subject*, by the last period of the no-punishment treatment of which all average group contributions converged to zero.

**Figure 5:** Comparison of Average Contributions in the P-experiment



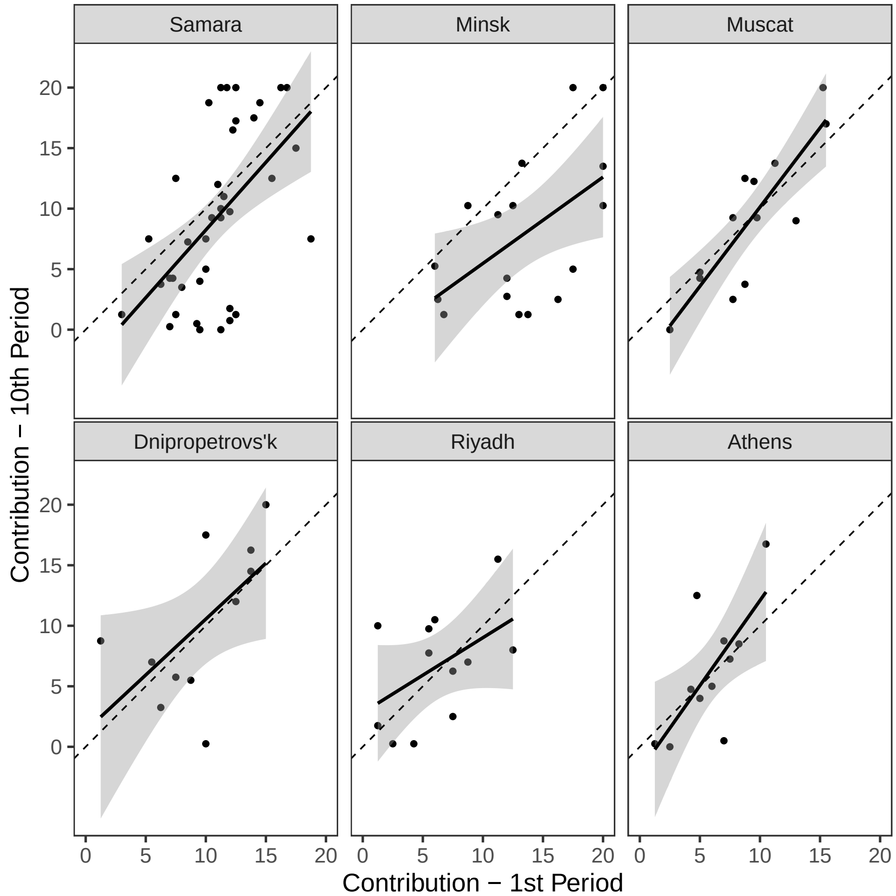

*Notes:* The horizontal axis denotes the 1st period mean contribution and the vertical axis denotes the 10th period average contribution in P-experiment. The unit of observation is the average group contribution. The points below dashed 45 degree line implies that those groups fail to improve upon the 1st period mean contribution while the points above the line show the groups that improve their mean contribution in the final period. The solid line denotes the linear fit while the shaded gray areas display 95% confidence intervals.

Istanbul is not the only city that exhibits sizeable heterogeneity in average first-period contributions that persist for the rest of the P-experiment. Herrmann et al. (2008) conduct their experiment in various cities with stark socio-economic differences. In order to explore which of these cities resemble Istanbul better, we conduct principal component analysis (PCA) using 1) straight-line physical distance to Istanbul 2) GDP per capita (in 2017 current US Dollars) 3) cultural and psychological distance to Istanbul (Turkey) via Muthukrishna (2018)'s WEIRD scale index. Our PCA exercise reveals that a cluster of six cities by Herrmann et al. (2008) resemble Istanbul the most: Athens (Greece), Dnipropetrovsk (Ukraine), Samara (Russia), Minsk (Belarus), Riyadh (Saudi Arabia), and Muscat (Oman).[11] We depict first-period average group contributions against their last-period counterparts via Herrmann et al. (2008)'s data for these six cities in Figure 6. We observe that in all of the six cities, there is sizeable first and last-period heterogeneity in average group contributions, along with a strong positive correlation between the two, which is close to unity. We further document that a (city) fixed-effect regression of last-period average group contributions on first-period average group contributions by these six cities yields a slope coefficient of 0.975 (with standard error 0.140), which is not statistically different than the slope of 0.997 (with standard error 0.243) that we estimate for Istanbul.[12] Accordingly, we conclude that Istanbul does not single out with its idiosyncrasy vis-à-vis heterogeneity in first-period average contributions and their persistence throughout the game. Instead, these stylized patterns are common across cities that share socio-economic similarities with Istanbul.

[11] Four of the cities by Herrmann et al. (2008) lack Muthukrishna (2018)'s WEIRD scale index scores. For the lacking cities, we imputed as follows: for Athens of Greece we proxied via Cyprus; for Riyadh of Saudi Arabia and Muscat of Oman, we proxied via neighboring Gulf countries, and for Copenhagen of Denmark, we proxied via Sweden and Norway. The resultant Figure A.2 is in the Appendix. Details of our PCA exercise are available upon request.

[12] The city fixed-effect regression of the six cities yields a constant coefficient of -1.139 with a standard error of 1.467 whereas the constant coefficient is 1.100 with a standard error of 2.022 for Istanbul. The respective Figure A.1 is in the Appendix.

**Figure 6:** First-Period and Last-Period Contributions by City (Herrmann et al., 2008)




*Notes:* The horizontal axis denotes the 1st period mean contribution and the vertical axis denotes the 10th period average contribution. The unit of observation is the average group contribution. The points below dashed 45 degree line implies that those groups fail to improve upon the 1st period mean contribution while the points above the line show the groups that improve their mean contribution in the final period. The solid line denotes the linear fit while the shaded gray areas display 95% confidence intervals.

# 6 Discussion and Concluding Remarks

In this paper, we report the results of an experiment conducted in Istanbul and demonstrate that in a public good game involving a punishment mechanism (P-experiment) contributions are significantly higher compared to a public good game without punishment (N-experiment), when a between-subject design is used. Contrary to this, Herrmann et al. (2008) find no difference between two conditions when a within-subject design is used and P-experiment follows N-experiment. However, the punishment mechanism is still not able to generate the average contribution levels observed in cities like Boston, Nottingham, Copenhagen, Bonn, Zurich, and St. Gallen. That is, the problem of cooperation seems persistent for public good games played in Istanbul. On the other hand, antisocial punishment is rare, has a spiteful motive behind it, and gets less and less common in later periods. Among the demographic and attitudinal variables, the response to the general risk question turns out to have a highly significant positive effect on contributions.[13]

A detailed look at the contribution levels of different groups in the P-experiment reveals that the cooperation problem manifests itself at very different levels for different groups. In the two panels of Figure 2, we provide a graphical representation of the evolution of contributions for all separate groups in the P-experiment. The left (right) panel of this figure involves groups that end up contributing less (more) than 50% of their endowment as of the final ($10^{\text{th}}$) period. As the figure reveals, the time path of contributions has a substantial variance among different groups in the P-experiment. Here, 4 out of 15 groups reach full contribution levels and 4 other groups reach contribution levels of 60% to 85% of the endowment. On the other hand, 4 of the remaining groups contribute around 40% of the endowment as of period 10 and the last 3 groups only contribute around 5%. The groups which start the game at a higher average contribution rate finish the game with higher average contributions, as well (right panel). As Sylwester et al. (2013) also discuss, some groups in Herrmann et al. (2008) immediately increase their contribution and continue to contribute high levels whereas some groups do not manage to increase their contribution much and contribute at relatively stable levels. This is reminiscent of what we observe for the subject pool in the current experiment. We can explain such differences in the evolution of contributions especially if the majority of the agents in a group are reciprocators (Burlando and Guala (2005)).

Our results for the P-experiment point out the importance of initial contributions to the subsequent evolution of contributions within a group. This result is reminiscent of previous studies demonstrating the effectiveness of grouping subjects based on their contributions to sustain high levels of cooperation. As an example, In Gunnthorsdottir et al. (2007), authors vary the marginal per-capita return (MPCR) and the assignment of subjects to groups. This assignment is done either using a random matching each period or through sorting of subjects in different groups according to the similarity of their contributions in the previous period. For each distinct MPCR value, net contributors exhibit a substantially lower decline in their contributions throughout the experiment when the latter type of matching is used. Studies reporting similar results include Gächter and Thöni (2005), Ones and Putterman (2007), Gunnthorsdottir et al. (2010). These studies demonstrate the effectiveness of grouping subjects based on their contributions to sustain high levels of cooperation. Using a similar approach, Brekke et al. (2011) allow subjects to choose between two groups, one of which requires a donation to a charity (Red-Cross) from the members. They find that those choosing the donation group are also more successful to sustain high levels of cooperation in

[13] The size of this effect is around 0.9 in model 2 and 0.2 in model 4, reported in Table 3.

a public good game experiment.

Gächter et al. (2010) mention two main channels where the culture can affect the behavior in dilemmas of cooperation. These channels are (i) beliefs of the subjects and (ii) the differences in responses to punishment. Beliefs matter mainly because subjects tend to cooperate more when they believe other group members will do the same (Croson, 2007; Gächter, 2007; Fischbacher and Gächter, 2010). Based on the results we report here, we believe that beliefs might have a stronger role, as evidenced by the high degree of the variance observed for initial contributions in the P-experiment. The fact that a within-subject design also induces lower contribution in the P-experiment (as in Herrmann et al., 2008) also supports this claim, since the subject's initial experience within the N-experiment might simply induce less optimistic beliefs in the P-experiment that follows. We conclude that beliefs about other's contributions and personal characteristics such as risk-taking highly influence the initial contributions to public good games. After this, contributions are adjusted based on the contributions of other group members and the punishment points received. Consequently, a low average contribution by the group at the beginning of the game and the existence of antisocial punishment might easily lead to failure to cooperate in this type of dilemma.

Herrmann et al. (2008)'s P-experiment data reveals that Istanbul is not unique in its first-period average contributions heterogeneity and its persistence over periods: cities that resemble Istanbul yield similar results. We show the prevalence of these stylized patterns in Athens (Greece), Dnipropetrovsk (Ukraine), Samara (Russia), Minsk (Belarus), Riyadh (Saudi Arabia), and Muscat (Oman). These cities were particularly chosen as they turned out to be the most similar ones to Istanbul in our principal component analysis. Accordingly, we conclude that concentrating only on city-level averages and ignoring within-city heterogeneity across groups result in misleading public good contribution inferences for several places around the globe.

Future research should focus more on the reasons for the variance in initial contributions in public good games and the role of culture in shaping these beliefs.

## References


Aghion, P., and Tirole, J. (1997). "Formal and Real Authority in Organizations". *Journal of Political Economy*, 105(1), 1–29.

Andreoni, J. and Miller, J. (2002). "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*, 70, 737–753.

Arifovic, J. and Ledyard, J. (2002). "Individual Evolutionary Learning, Other-Regarding Preferences, and the Voluntary Contributions Mechanism." *Journal of Public Economics*, 96, 808–823.

Brekke, K. A., Hauge, K. E., Lind, J. T., & Nyborg, K. (2011). "Playing with the Good Guys. A Public Good Game with Endogenous Group Formation". *Journal of Public Economics*, 95(9-10), 1111–1118.

Burlando, R., and Guala, F. (2005). "Heterogeneous Agents in Public Goods Experiment". *Experimental Economics*, 8(1), 35-54.

Cohen, J. (1988). "Statistical Power Analysis for the Behavioral Sciences". Erlbaum, Hillsdale, NJ.

Croson, R. T. (2007). "Theories of Commitment, Altruism and Reciprocity: Evidence from Linear Public Goods Games". *Economic Inquiry*, 45(2), 199–216.

Fehr, E., and Gächter, S. (2000). "Cooperation and Punishment in Public Goods Experiments". *American Economic Review*, 980–994.

Fehr, E., and Gächter, S. (2002). "Altruistic Punishment in Humans". *Nature*, 415, 137-?140.

Fischbacher, U., Gächter, S. (2010). "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments". *American Economic Review*, 100(1), 541–556.

Gächter, Simon. 2007. "Conditional Cooperation: Behavioral Regularities from the Lab and the Field and Their Policy Implications". In Psychology and Economics, Bruno S. Frey and Alois Stutzer, 19–50. Cambridge: The MIT Press.

Gächter, S., Herrmann, B., and Thöni, C. (2010). "Culture and Cooperation". *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365(1553), 2651–2661.

Gächter, S., Thöni, C., 2005. "Social Learning and Voluntary Cooperation among Likeminded People". *Journal of the European Economic Association*, 3 (2–3), 303–314.

Gunnthorsdottir, A., Houser, D., and McCabe, K., and McCabe, K. (2007). "Disposition, history and contributions in public goods experiments". *Journal of Economic Behavior & Organization*, 62(2), 304–315.

Gunnthorsdottir, A., Vragov, R., Seifert, S., and McCabe, K. (2010). "Near-Efficient Equilibria in Contribution-Based Competitive Grouping". *Journal of Public Economics*, 94(11-12), 987–994.

Herrmann, B., Thöni, C., Gächter, S. (2008). "Antisocial Punishment Across Societies". *Science*, 319(5868), 1362–1367.

Isaac, R. M., & Walker, J. M. (1988). "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism". *Quarterly Journal of Economics*, 103(1), 179–199.

Lee-Panagos, A. (2016). "Modelling Contributions In Public Good Games With Punishment". CeDEx Discussion Paper Series No. 2016-15.

Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2018). "Beyond WEIRD Psychology: Measuring and Mapping Scales of Cultural and Psychological Distance." SSRN Working Paper No: 3259613.

Ones, U., Putterman, L., 2007. "The Ecology of Collective Action: a Public Goods and Sanctions Experiment with Controlled Group Formation". *Journal of Economic Behavior and Organization*, 62 (4), 495–521.

Ostrom, E., Walker, J., and Gardner, R. (1992). "Covenants with and without a Sword: Self-Governance is Possible". *American Political Science Review*, 86(02), 404–417.

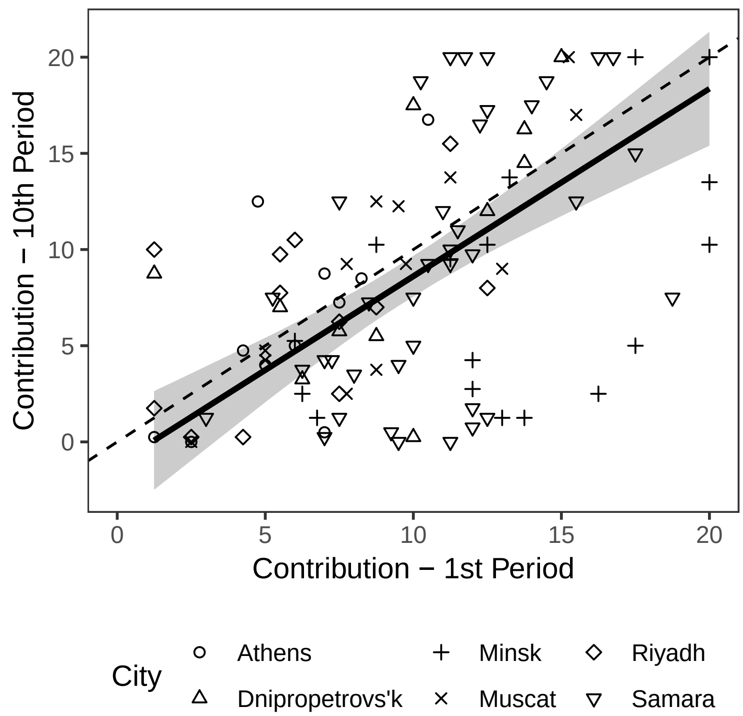
Sylwester, K., Herrmann, B., and Bryson, J. J. (2013). "Homo Homini Lupus? Explaining Antisocial Punishment". *Journal of Neuroscience, Psychology, and Economics*, 6(3), 167.

Yamagishi, T. (1986). "The Provision of a Sanctioning System as a Public Good". *Journal of Personality and Social Psychology*, 51(1), 110.

# Appendix

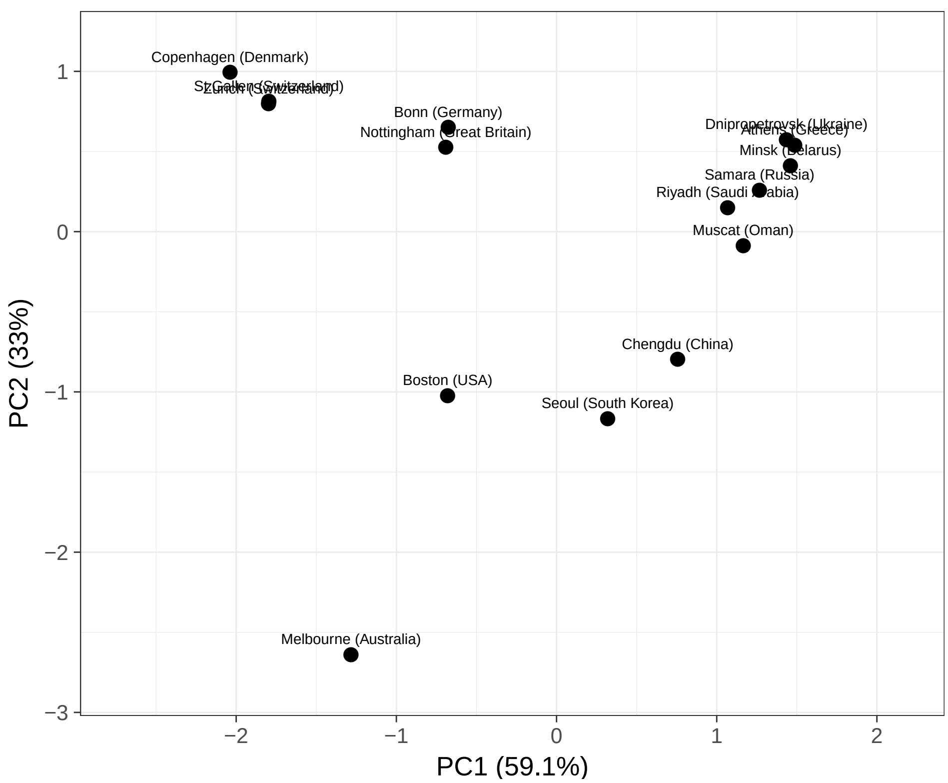
## A Appendix Figures

**Figure A.1:** First-Period and Last-Period Contributions by City (FE) (Herrmann et al., 2008)




*Notes:* The horizontal axis denotes the 1st period mean contribution and the vertical axis denotes the 10th period average contribution in 6 cities resembling to Istanbul from PCA analysis. The unit of observation is the average group contribution. The points below dashed 45 degree line implies that those groups fail to improve upon the 1st period mean contribution while the points above the line show the groups that improve their mean contribution in the final period. The solid line denotes the linear fit while the shaded gray areas display 95% confidence intervals.

**Figure A.2:** Principle Component Anaylsis (Herrmann et al., 2008)




*Notes:* Figure A.2 displays the principle component analysis scores of the fifteen cities by Herrmann et al., 2008 via 1) straight-line physical distance to Istanbul 2) GDP per capita (in 2017 current US Dollars) 3) cultural and psychological distance to Istanbul (Turkey) via Muthukrishna (2018)'s WEIRD scale index.