Prosocial Behavior and the Individual Normative Standard of Fairness within a Dynamic Context: Experimental Evidence

Mekvabishvili, Rati and Mekvabishvili, Elguja and Natsvaladze, Marine and Sirbiladze, Rusudan and Mzhavanadze, Giorgi and Deisadze, Salome

Ivane Javakhishvili Tbilisi State University

4 February 2023

Prosocial Behavior and the Individual Normative Standard of Fairness within a Dynamic Context: Experimental Evidence

Rati Mekvabishvili[1]*, Elguja Mekvabishvili[1], Marine Natsvaladze[1], Rusudan Sirbiladze[1], Giorgi Mzhavanadze[1], Salome Deisadze[1]

Published in Open Journal of Social Sciences, Vol.11 No.3, March 2023

Abstract

In this paper, we present an experimental study of prosocial behavior and individual normative standards of fairness under the novel context of a dynamic dictator game. In addition, we explore the role of informal institutions in shaping individuals' cooperation within the domain of a public goods game under its direct exposure and in subsequent prosociality beyond its reach in the domain of the dictator game. We find that dictators' average offers in our study are quite close to the typical results found in other dictator game experiments and they are quite stable over two periods. However, dictators become more selfish after they have had the experience of playing a public goods game with peer punishment. Interestingly, we found that dictators act significantly more selfishly relative to their own declared individual normative standard of fairness. Furthermore, our experiment reveals a large share of antisocial punishment in the public goods game and a peer-to-peer punishment mechanism to be an inefficient tool to promote cooperation, however in an environment that rules out a suitable normative consensus and collective choice.

JEL-Code: D02, C73, C92, H41
Keywords: dictator game, individual normative standard of fairness, dynamics of behavior, spillover, prosociality, public goods game

[1] Faculty of Economics and Business, Ivane Javakhishvili Tbilisi State University, University st.N2, Post box 0186, Tbilisi, Georgia.
*Corresponding author: rati.mekvabishvili@eab.tsu.edu.ge , ORCID: 0000-0002-9027-1802
This research was funded by the Faculty of Economics and Business at Ivane Javakhishvili Tbilisi State University. We thank Deyan Dzhurov for his valuable assistance in programming the experiment. We are grateful to Benedikt Herrmann, Nick Stagnaro and Antonio Cabrales Goitia for their valuable comments and suggestions.

1. Introduction

The subject of motivation for the engagement of purely altruistic and fair behavior has been of research interest for many economists and social scientists. More importantly, further questions still surround the type of institutional incentives that promote prosocial behavior and the normative underpinnings behind such behaviors. An increasing body of research has been devoted to the importance of social preferences during decision-making in various contexts. The study of social preferences has also been one of the most significant topics for experimental economists over the last three decades.

Our paper contributes to the growing body of literature on prosocial behavior and norms of fairness with several distinct extensions. Firstly, we investigate the stability of prosocial behavior and whether immediate repetition of identical situations has an impact on such behavior or on fairness. For instance, does behavior remain stable if decisions concerning one's selfishness or fairness are repeated? Social dilemmas are typically not one-time encounters, rather repeated games and raise the question of whether behaviors change due to repetition. Equally, when someone is exposed to two identical situations over a given period, they might also encounter certain other experiences during the intervening stage. Indeed, most real-life situations in which people have to decide whether to be selfish or to behave altruistically are not single events that arise only once in a lifetime. It is therefore important to know more about the dynamics of behaviors driven by social preference. Since the dictator game (DG) is thought to measure pure altruistic and prosocial behavior, we conduct a two-round, double-blind standard dictator game experiment. It is well-established in the literature that the dictator game is a suitable tool for measuring altruism and prosociality since subjects freely (and anonymously) decide their level of donation, where tension can arise between selfishness and fairness.

Secondly, we examine individual normative standards of fairness. Generally, the problem with norms is that they are difficult to measure, yet Camerer and Fehr (2004) discuss a variety of instruments used to measure social norms. In order to evaluate the individual normative standard of fairness, we apply a standard dictator game – explicable as the game is also used to measure fairness in the allocation of resources. Within the game, participants are free of any strategic considerations and their behavior is predominantly motivated by their altruistic considerations, as well as by norms of what is regarded as fair and what is not. In particular, we elicit individual normative standards of fairness by applying the norm elicitation method introduced by Fehr and Williams (2017) to learn more about underlying norms in repeated settings. Therefore, in both periods of the dictator game before the decision stage, the subjects were asked to answer a specific normative question. This design setup additionally enabled us to measure the norm of fairness in a dynamic context. The question we wish to address is what the individual normative standard of fairness denotes regarding the appropriateness of behavior in repeated situations. For example, do the subjects demand that the same transfer always has to be made? Moreover, does the actual distribution of offers from dictators correspond to their individual normative standards of fairness?

Finally, our paper focuses on whether informal (peer-to-peer) institutional incentives to cooperate can have an impact on prosociality, and particularly if prosociality is influenced in distinct situations beyond the reach of such an institution. More specifically, we examine a particular spillover-based

theory that is more explicitly cognitive than most: the Social Heuristics Hypothesis (SHH) (Bear and Rand (2016), Rand et al., (2014)). We are interested in the impact of a public goods game with peer-to-peer punishment mechanism on cooperative behavior – under its direct exposure as well as in the subsequent behavior after its removal. For this purpose, we constructed an experiment design consisting of three stages. In the first stage, we conducted a standard double-blind dictator game; in the second stage there was a ten-round public goods game (PGG) with peer punishment; and in the third, final stage, the subjects played the same dictator game. Therefore, by comparing the behavior of dictators in the first and final stages, we can analyze the impact of the peer-to-peer punishment mechanism in the public goods game on subsequent prosocial behavior. In addition, these details can thereafter be compared to our control treatment, where the subjects play an identical two-round dictator game.

The remainder of the paper is organized as follows: Section 2 reviews the related literature; in Section 3 we describe the experimental design; Section 4 provides the experimental results, which are then discussed in section 5; and Section 6 offers the respective conclusions.

2. Related literature

Our study is related to several aspects of the experimental literature. First, the work relates to the literature that has examined the role of formal institutions in shaping individuals' cooperation and subsequent prosociality. We focus here on experimental studies of a particular spillover-based theory (Bear and Rand (2016), Rand et al., (2014)). Peysakhovich and Rand (2016) in their experimental study examined the link between peer-based reputational incentives in cooperation and subsequent prosociality, and they experimentally demonstrated the spillovers of prosociality. Their study posed a two-stage experiment, wherein first stage subjects play repeated prisoner's dilemmas in conditions that do or do not support cooperation, and in the second stage they play a one-shot dictator game. They found that subjects from environments that support cooperation are more prosocial.

Further empirical support for the spillover-based theory comes from evidence that experimentally manipulated institutional strength, within a public goods game with a centralized punishment mechanism, led to significantly more prosociality in a subsequent one-shot dictator game (Stagnaro et al., 2017). In their study, the emphasis was on the finding that strong formal institutions which incentivize prosociality also positively affect prosociality. Namely, the motivation to cooperate in the public goods game domain increased subsequent giving in the domain of the dictator game. A recent PGG experiment by Engel et al., (2021) study how the presence and nature of an exogenously and endogenously imposed institution that enforces prosocial behavior in one domain affects behavior in another domain. They found clear evidence in favor of such positive spillover effects. Mekvabishvili (2021) studied the spillover effect on cooperation in a single domain, as measured by a repeated anonymous public goods game with an exogenous centralized punishment mechanism with probability. Specifically, they conducted a two-stage experiment, where during the first stage the subjects played a PGG with exogenous centralized punishments, with probabilities for ten rounds, while in the second stage, the punishment mechanism was removed and the subjects played ten rounds of a standard PGG without punishment. Their findings do not provide support for the spillover-based

theory. Exposure to formal institutions that provide top-down motivation for cooperation substantially improves cooperation in their presence, however it does not seem to instigate more prosociality after their absence. Here, we question whether the positive connection between centralized punishment-based formal institution incentives to cooperate in one domain and the subsequent prosociality in another domain, as found by Stagnaro et al., (2017), extends to a peer-to-peer punishment-based informal institution.

Second, our paper is related to the experimental studies on the subject of behavior in a dictator game, alongside norms of fairness in a dynamic context. The dictator game (Kahneman et al., (1986), Forsythe et al., (1994)) is the most basic decision situation in which social preferences can be studied. The dictator game itself is a two-player game where participants are randomly assigned to be either a "dictator" or a recipient. The dictator has to decide how much of a given amount of money is allocated to themselves or shared with the recipient, who has to accept the offer. Usually, equal division of the monetary rewards or costs is a widely observed behavioral norm, and compliance with the 50/50 split in dictator games has been well documented in laboratory experiments. Based on the definition of altruistic behavior (i.e., more cost to oneself and more benefit for others), a smaller than equal division is not considered altruistic and only a division equal or larger than 50% would be regarded as altruistic behavior. In dictator game experiments in which the equal split norm prevails, around 30% of dictators send half of their endowment to the recipients Camerer (1997). Thus, the results of many carefully controlled dictator games do not support self-interest predictions on average. Furthermore, a meta-analysis Engel (2011) finds that across 616 treatments involving the dictator game, the average sharing rate is 28.3 percent across all studies, with about 36 percent of individuals not sharing at all. As such, many people are willing to share a windfall gain, yet a considerable minority are not. In addition, this meta-study on the dictator game covering 129 papers, finds that, when the game is played repeatedly, dictators offer lower amounts and equal splits become less likely.

Surprisingly, the literature is rather mute regarding the dynamic aspect of the dictator game. The stability of prosocial behavior was investigated for the first time by Brosig et al., (2007), who conducted a modified series of repeated dictator experiments. During which there was a time span of four weeks between the repeated experiments. They ultimately observed that the behavioral dynamics had only a single direction: from less selfish to more selfish behavior. In later experiments, Sass and Weimann (2015a) investigated the behavioral dynamics in repeated trust and mutual gift-giving games, and they also found that the propensity to give decreases after repetition of the games. Sass and Weimann (2015b) again reported the same during a series of repeated standard public good games. While Sass et al. (2015c) examined how social distance influences behavior in repeated dictator experiments across different time spans, and they found that behavioral dynamics shift from less selfishness to greater levels of selfishness. They equally discerned that overall altruistic giving decreases over time and that this decline does not depend on the time span between repetitions. In addition, they used the method developed by Krupka and Weber (2013) to elicit the norm of cooperation. They found that feeling approved to behave more selfishly after being generous seems to be covered by social approval. Brosig et al. (2017) investigate the dynamics of individual pro-social behavior over time in dictator game. The dynamics are tested by running the same experiment with the same subjects at several points in time. They found that prosocial decisions decrease over time. As for the stability of behavior in the sense that subjects stick to their decisions over time is observed predominantly for purely selfish subjects.

Thirdly, our work is connected to the experimental literature on social norms. The method for identification of social norms in social dilemma games, as applied here, has recently been presented in two papers (Fehr and Williams (2017), Fehr and Schurtenberger (2018)). In the public goods game, each group member is asked to indicate what other group members should contribute to the public good. Thus, by answering this normative question, subjects conveyed their typical standards of cooperation. Thereafter, subjects' average normative requests are conveyed to all group members and then likely constitute a general standard of cooperation – one which is commonly known and reflects the views of group members. Moreover, the higher the subjects' agreement with their normative requests, the more the average request will constitute a legitimate normative standard. This approach enables the identification of an individual normative standard and a normative consensus among group members and also leads to norm formation. Fehr and Williams (2017) in their experimental study identified that efficient peer sanctioning without great need for costly punishment emerges quickly when subjects have the chance to achieve a consensus about normatively appropriate behavior. Therefore, the existence of a normative consensus is critical for cooperation to flourish.

3. Experimental design

3.1 Participants

The experiment was conducted in Georgia via the Lioness software platform, used for interactive online experiments Arechar et al., (2018). A total of 146 subjects participated, mostly from Tbilisi State University. The experimental data are available at the Zenodo data repository Mekvabishvili et al. (2022). The average age of participants in the sample was 20.6 years and 63.8% were female. The participants majored in various subjects: 39.9% were students of Economics, 14% in Business Administration, humanities accounted for 5.5%, law 5%, medicine 4.4%, other majors 14.9%, and 16.3% were non-students.

In the control treatment (CT), the task took participants between 10 and 15 minutes to complete, and their earrings on average totaled 10.9 GEL (4.0 USD at the time, min 0.0/max 7.5). In treatment 1 (T1), the experiment session lasted between 20 and 25 minutes and the participants earned on average 15.5 GEL (5.5 USD at the time, min 0.0/max 15.0). After finalizing the experiment, as soon as the participant provided an electronically signed payment document, they were immediately paid via internet bank transfer.

3.2 Method

In total, six experimental sessions were conducted. Our experiment consisted of two treatments: the control treatment (CT) and treatment 1 (T1). We prevented repeated participation by excluding duplicated IDs and IP addresses. The participants were not informed about the identity of their group members. In the CT, the subjects were randomly assigned in pairs to play a two-round, double-blind anonymous dictator game either as the dictator or the receiver. A subject in the role of dictator in each round is endowed with certain amount of points and has to decide how much to give to another

subject. A subject in the role of receiver has to accept the dictator's offer. In both rounds, we kept paired subjects constant. T1 consisted of three stages. In the first stage, the subjects played a standard, double-blind, one-round dictator game. During the second stage, the participants completed ten periods of a public goods game with peer punishment. At the beginning of stage 2, the participants were randomly selected in groups of four and the group members were then constant across all ten periods. In the third stage, the players again participated in the same one-round dictator game. Critically, the roles in the dictator games were all fixed. Thus, in stage 3 when the subject played the same dictator game again, the dictators were the same subjects as in stage 1. However, during the dictator game, the subjects were randomly matched in both stages, and the participants were informed about matching. This experimental setup enabled us to track the same dictators' decisions over time. Moreover, applying the method elaborated by Fehr and Williams (2017), we can elicit individual normative standards of sharing in the control treatment; during each period before the decision part of the experiment, the subjects were asked to provide their individual normative standard of sharing.

3.3 Information conditions

To maximize data quality, we required game comprehension before commencing both treatments: after reading the instructions, the participants could not proceed to the game until they correctly answered all control questions (they were allowed an unlimited number of attempts). To ensure that participants did not have varying expectations about the length of the experiment, the total number of rounds/stages was public knowledge in both treatments. In the CT, detailed instructions were given to participants, and both players could observe the results after the division of points in each period. However, they could not observe their normative standards, as declared in the norm formation part. During treatment 1, more detailed instructions about the game were given before the start of each stage. To avoid issues related to potential income effects in T1, we followed a common practice in experimental economics and informed participants that only one of the three stages would be randomly selected for payment (so that earning more points in one stage would not make participants feel like they had more to spend in subsequent stages). In stage 3, the subjects were informed that the same game was starting as in stage 1, although they were paired randomly with a new player.

In stage 2 of treatment 1, the subjects played ten rounds of a standard linear public goods game with peer punishment, as developed by Fehr and Gachter (2000). They were informed that they were grouped randomly with new subjects. They were also notified that the composition of each group would remain unchanged throughout stage 2. In each round, the players made their contribution decisions simultaneously and once the decisions were made, they were informed about their group members' contributions. Subsequently, these subjects entered the second round, where they could assign up to ten punishment points to each group member. Punishment was costly for both punished and punishing subject, where each deduction point assigned reduced the punished member's earnings by three points and the punishing member's earing by one point. All punishment decisions were made simultaneously. The participants were not informed about who had punished them.

3.4 Payoffs

In the control treatment, the dictators were endowed with 200 points in each round and had to decide how to allocate this amount between themselves and the receiver. In stage 1 and stage 3 of T1, the subjects played the same standard dictator game as in the CT, and they were endowed with 600 points. Table 1 depicts the payoff in the dictator game during each treatment.

Table 1: Dictator game in the CT and T1 treatments

	Payoffs	
Treatment	Dictator	Receiver
CT	200-x	x
T1	600-x	x

In stage 2 of the CT, the subjects were grouped into members of four, who played a public goods game with peer punishment. In every period, each member of the group received an endowment of 20 points. The participants had to decide how many points to keep for themselves and how many to contribute to a group project.

In the standard linear PGG over ten periods, the subjects simultaneously decided how much of the 20 endowment points to keep or invest in public goods during each period. The payoff is determined by $\pi_i^1 = 20 - g_i + 0.375 \sum_{j=1}^{n} g_i$; where g_i is a subject's contribution to public goods, and 0.375 is the marginal per-capita return of contributing to public goods. The total *i* is the sum of the period payoffs over all ten periods. Because the cost of contributing to the project was exactly one point, while the return on that point was only 0.375 points, retaining all of one's points was always in any participant's material self-interest, irrespective of how much the other three group members contributed. Thus, if each group member retained all their points, there were no earnings to be shared. On the other hand, every member would earn 0.375 × 80 = 30 points if each of them invested their entire endowment.

In each round, as soon as the subjects were informed about the contribution of the other three group members, a second part followed. The subjects could then punish their group members by assigning so-called "deduction points". The allocation of deduction points (p_{ij}), by player *i* to player *j*, reduces the first-part payoff of player *i* by one point and that of player *j* by 3 points. If player *i* receives p_{ij} deduction points from the other group members and assigns p_{ij} deduction points to member *j*, the final payoff of subject *i*, π_i, is:

$$\pi_i = \pi_i^1 - (3 \sum_{j=1}^{n} p_{ji} + \sum_{j=1}^{n} p_{ij}),$$

Consequently, a punishment decision was implemented by assigning the punished member between zero and 10 deduction points. Each deduction point assigned reduced the punished member's earnings by three points and cost the punishing member one point.

4. Results

The following analyses focuses on several sets of questions. Firstly, we are interested in whether there were differences in dictators' behavior during T1 compared to that in the CT. Secondly, we would like to distinguish whether the repetition of identical situations has an impact on prosocial behavior. Generally, one has to decide repeatedly when giving something in order to help others or just to be nice to others. The time span between such situations may vary, but we have analyzed it in the duration of a single experimental session. Thirdly, we would like to examine the effectiveness of the peer punishment mechanism in fostering cooperation in the PGG. Additionally, we explore whether the peer punishment mechanism, as an informal institution that incentivizes cooperation, will promote prosociality via spillovers, as predicted by spillover-based theory. Here, we address this issue by examining the relationship between an institution and one's prosociality (as measured by the dictator game).

4.1 Analysis of players' offers in the dictator game

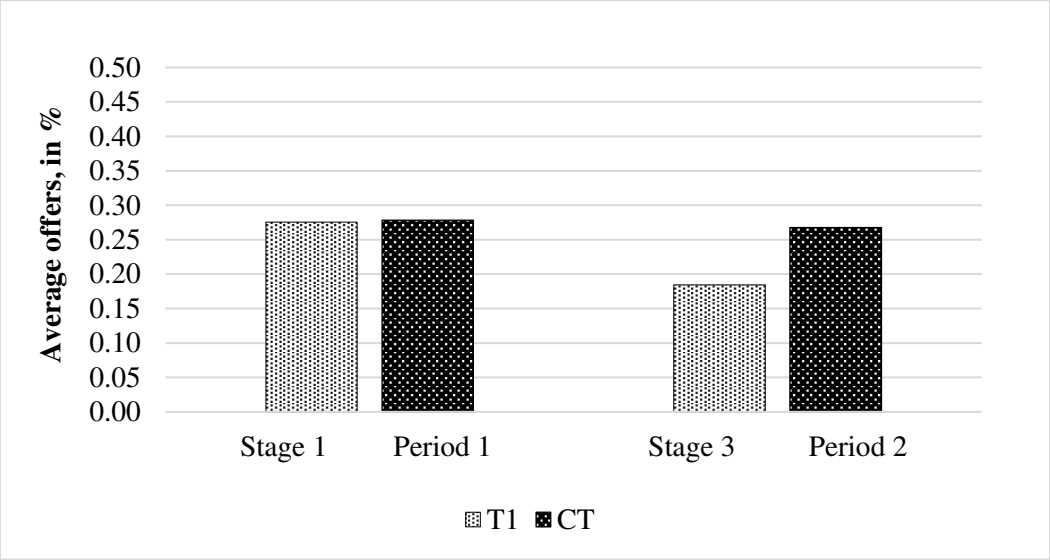
We begin with offers in the dictator games from both treatments. In this regard, we started with an analysis of subjects' offers within the treatments and then across treatments. We first analyzed participants' behaviors when playing the role of dictator in the control treatment (CT) and then compared it to dictators' behavior in the corresponding treatment (T1). Table 2 shows dictators' offers in both treatments. The dictators' average offers in our study are fairly similar to the results of a meta-analysis from Engel (2011), which found that across 616 treatments involving the dictator game, the average offer was 28.3 percent. In T1 and CT, on average, 36% and 23% of dictators respectively behaved fully selfishly (offered zero points), while 22.2% and 31.1% of dictators on average behaved altruistically, respectively (offering equal or more of their endowment). The descriptive analyses regarding dictator game behavior reveals that average offers from dictators remained relatively stable in the control treatment.

Table 2: Comparison of dictators' offers in CT and T1 treatments

N. of dictators who offer	**CT (n=37)**		**T1 (n=36)**	
	Period 1	**Period 2**	**Stage 1**	**Stage 3**
>50%	3	2	4	1
=50%	9	10	7	4
<50%	25	25	25	31
Average offer (in %)	27.8%	26.8%	27.5%	18.4%
Standard deviation (in %)	24.7%	23.0%	30.0%	22.0%

Although average dictators' offers are quite similar to those in the CT during the first stage of T1, one can notice a profound decrease in these average offers during stage 3 (see Figure 1 below). Interestingly, dictators behaved more selfishly when the dictator game was played after the PGG with peer punishment.

Figure 1: Average offers of dictators in CT and T1 treatments



Thus, playing ten periods of the PGG with peer punishment affected dictators' average offers and it increased the number of uneven split offers in the dictator game. Nevertheless, the described effect was not statistically significant (Mann-Whitney Test, p = 0.3162, z = -1.0023). Thereafter, we compared dictators' offers from both periods of the control treatment. The corresponding comparison of proposers' demands revealed that this difference was not statistically significant either (Mann-Whitney Test, p = 0.8120, z = -0.2378). The decreasing average offers of dictators in stage 3 relative to stage 1 suggests that exposure to informal institutions that motivate cooperation seems to lead to more selfishness rather to more prosociality after its absence.

4.2 Norm elicitation

Studying normative issues, one requires a clear definition of social norms. Here, we use the norm classification provided by Fehr and Williams (2017). They define norms as commonly known standards of behavior that are based on widely shared views of how an individual should behave in a given situation. In our experimental study, we apply individual normative standard elicitation method to learn more about the underlying normative standard in repeated settings. The basic notion behind their approach is to gather individual normative standards that rely on subjects' period-by-period normative requests. We applied their method in our control treatment with two-round dictator game. Namely, in both periods before each decision stage the subjects were asked to answer the following question: *"In your opinion, how should be the endowment points be divided between you and another participant?"*. The subjects only needed to insert a single number to answer this question. The participants thus reveal their normative standard of fairness by answering this question, and this enables us to measure their normative standard over time and to compare it to their actual sharing decisions.

Figure 2: Comparison of mean individual normative standards and offers of dictators

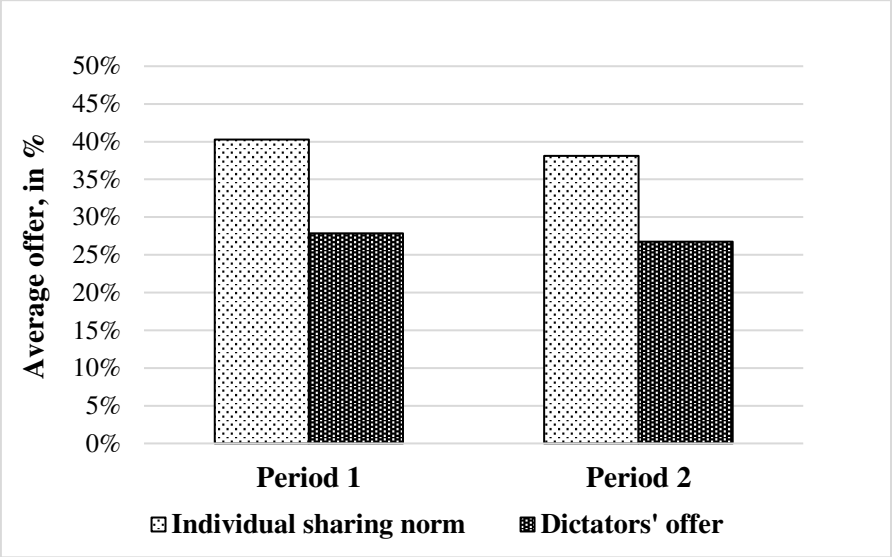
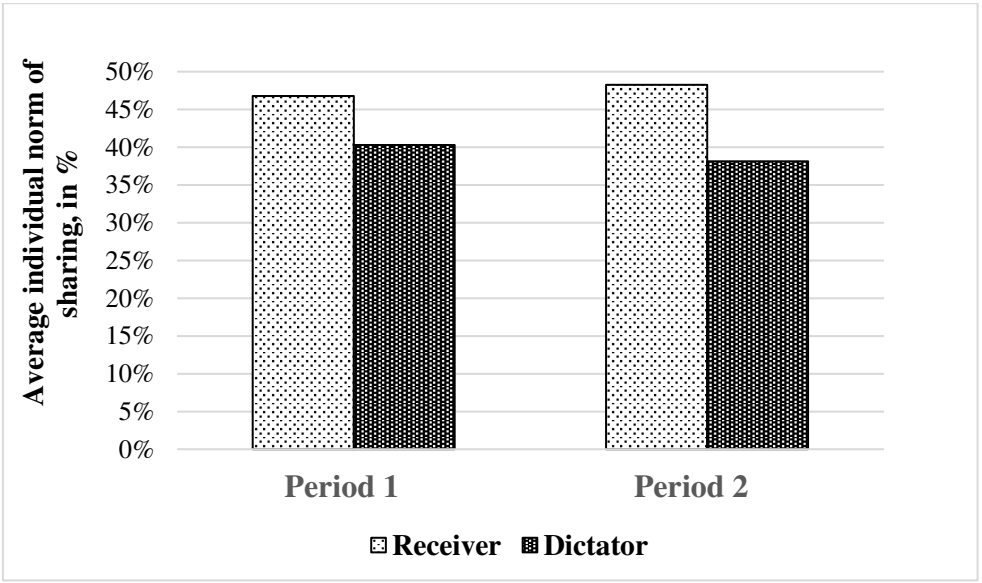


Figure 2 reveals that the opportunity to indicate an individual normative standard has no impact on behavior. This is likely due to our design setup, since the normative standards declared by the subjects were not a public information. In our case, we intended to keep these individual standards of fairness anonymous in order to uncover the normative motives in a more isolated condition. Interestingly, Figure 2 shows that dictators' normative standards are closer to an equal split and significantly higher than the actual offers proposed to the recipients in both periods (period 1: Mann-Whitney Test, p = 0.0323, z = -2.1405; and period 2: Mann-Whitney Test, p = 0.0263, z = -2.2216). The individual normative standard of fairness is relatively stable and statistically indistinguishable during both periods (Mann-Whitney Test, p = 0.8120, z = -0.2378). There are substantial deviation of actual offers from individual normative standard of fairness. As such, the existence of a relatively high individual normative standard of fairness and prosociality is, per se, does not necessarily converts in the same level of prosocial behavior, suggesting that intrinsic motives for an individual normative standard of fairness compliance are not sufficiently strong. This is in line with the findings of Fehr and Schurtenberger (2018).

Figure 3: Mean individual normative standard of sharing in CT



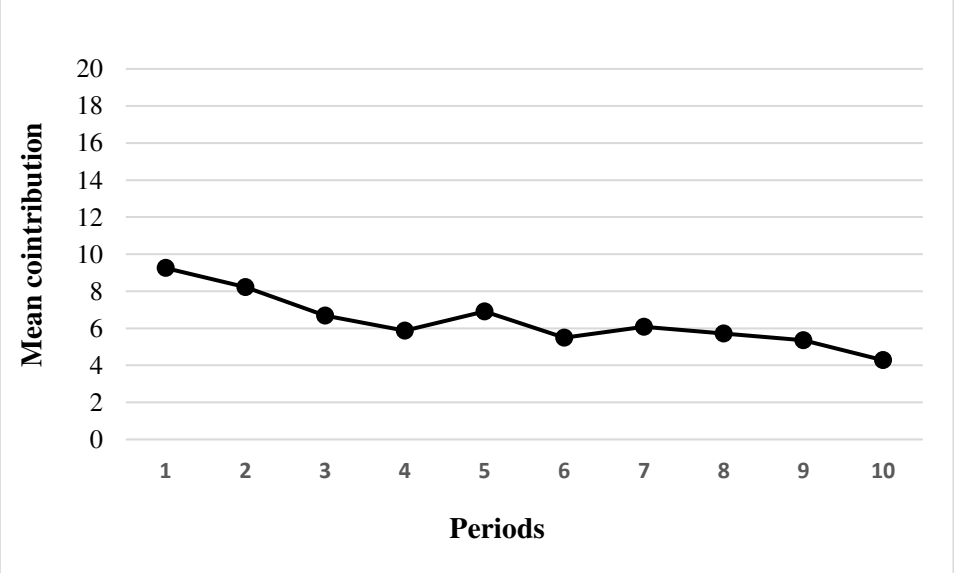
Next, we examined the individual normative standards of fairness for recipients and compared it to that of dictators (Figure 3 depicts the results). Recipients' individual normative standard of fairness is closer to an equal split and higher than those declared by dictators, although the difference is not statistically significant (period 1: Mann-Whitney Test, p = 0.3638, z = -1.9081; and period 2: Mann-Whitney Test, p = 0.0917, z = -1.0865). This observation suggests that in the absence of a sanction mechanism or recipient's bargaining power, individual normative standard compliance is disregarded to a considerable degree.

4.3 Cooperation levels in PGG with peer punishment

The efficacy of peer-to-peer punishments in improving cooperation is well-reported (Fehr and Gächter (2000), Fehr and Gächter (2002), Nikiforakis (2008), Gächter et al., (2010)). Although, the existence of punishment opportunities in public goods games causes strong increases in cooperation for Western, but not for all, cultures Herrmann et al., (2008). In particular, for countries with weak norms of civic cooperation, the antisocial punishment of cooperators is particularly strong, and it is associated with detrimental effects on overall cooperation rates.

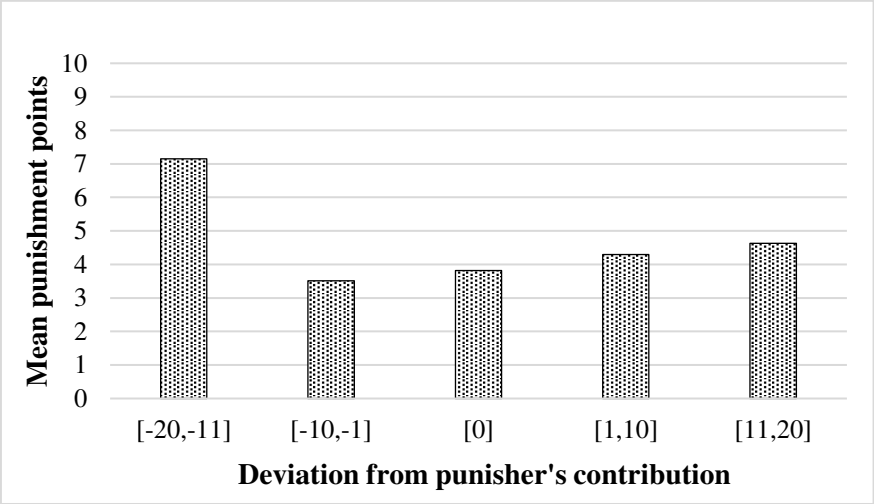
We began our analysis by assessing how PGG contributions developed across the periods under the peer punishment mechanism in stage 2 of T1. Figure 4 shows average contributions over time. In Figure 4, the mean contributions exhibit a declining pattern and the contributions across all rounds remain quite low – well below 50%. The mean contributions averaged at 6.4 points (standard deviation 1.46), which is 32% of the endowment. In the last period, 58 percent of the subjects contributed zero, while in the first period only 32%; the difference between the contributions during these two periods is statistically significant (Wilcoxon signed rank test, two-sided, p = 0.000). Therefore, from Figure 4 one can identify that the introduction of a peer punishment mechanism appears to be an ineffective tool to promote cooperation and discipline free riders.

Figure 4: Mean contributions in PGG with peer punishment in stage 2



Next, we turn to punishment behavior. Namely, we investigated how an individual who has contributed a certain amount to the public good then punishes other group members who contributed either less, the same amount, or more than their own contribution. Figure 5 displays punishment expenditures as a function of how much a punished individual's contribution deviated from the contribution of the punisher.

Figure 5: Mean punishment expenditures



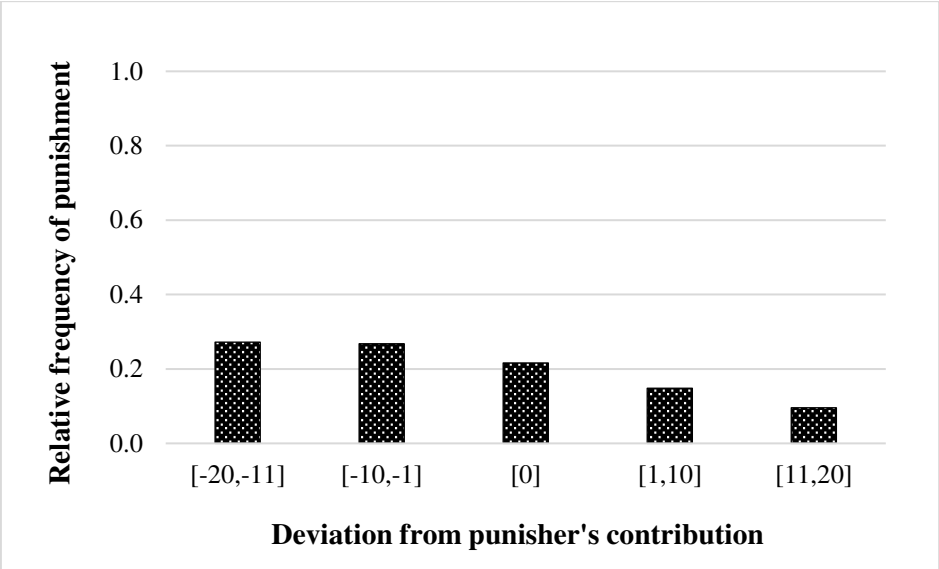
In Figure 5, the various deviations of the punished participant's contribution from the punisher's contribution are grouped into five intervals.[2] From Figure 5 one can thus observe that those punished subjects who contributed between 11 and 20 points less than punishing subjects, on average, received 7.1 punishment points from the punishing subjects, and 4.6 points on average in the case of positive deviations between [11, 20]. According to the experimental study by Herrmann et al. (2008), the punishment for negative deviations is labeled as a penalty for free riding; and the punishment for positive deviations is regarded as antisocial punishment since those who were penalized contributed either the same or even more than those who were punishing. As highlighted in Figure 5, although free riders are punished strongly, there remains a substantial share of antisocial punishment.

Figure 6 depicts the relative frequency of punishment for a given deviation from the punisher's contribution. Deviations of the punished subject's contribution from the punisher's contribution are grouped into the same five intervals. From Figure 6 we can observe that the relative frequency of punishment for a given deviation from the punisher's contribution across five intervals does not vary substantially. This suggests that no common behavioral standard for individual contributions is emerging. As in Figure 5, we find that the probabilities for punishing a free rider and antisocial punishment are quite similar, at 54% and 46%, respectively. These observations indicate that

[2] The interval [–20, –11] indicates that the punished participant contributed between 11 and 20 tokens less than the punishing participant; [–10, –1] indicates that the punished participant contributed between 1 and 10 tokens less than the punishing participant; [0] indicates that the punished participant contributed the same amount as the punishing participant; [1, 10] indicates that the punished participant contributed between 1 and 10 tokens more than the punishing participant; and [11, 20] indicates that the punished participant contributed between 11 and 20 tokens more than the punishing participant.

punishment had a fairly weak disciplinary effect on free riders, in the sense of steering low contributors toward higher contributions.

Figure 6: Relative frequency of punishment



Our results show that in the CT average offers from dictators remained relatively stable. However, in the T1 treatment playing ten periods of the PGG with peer punishment led to more selfish behavior of the dictators. Interestingly, declared level of individual normative standard of fairness of the dictators does not match with their actual giving, their offers are less fair than their individual normative standard of farness. Next, we found substantial level of antisocial punishment in PGG and that peer-to-peer punishment turned to be an ineffective tool to discipline free riders.

5. Discussion

We begin our discussion by considering the prosocial and altruistic behaviors in dictator games throughout both treatments, CT and T1. The results of our dictator games do not support self-interest predictions, where many people are willing to share a windfall gain. On the whole, our results on the average sharing rate are fairly typical, and close to the findings of a large meta-study Engel (2011) across 616 treatments involving a dictator game. The CT treatment demonstrates that, in a standard dictator game experiment, altruistic behavior is quite stable over two rounds.

Exploring of individual normative standard of fairness reveals that declaring higher standards of fairness, but actually offering substantially lower, is widely accepted. Moreover, we find that recipients' individual normative standards of sharing are higher and are closer to an equal split than dictators, which seems to reflect that recipients are powerless and that dictators fully exercise their bargaining power. This additionally suggests that an individual normative standard of fairness, if not backed by an enforcement mechanism or punishment, is followed only loosely. We believe that discrepancy between individual normative standards of fairness and actual offers of dictators is an interesting study for future research.

Our experimental evidence demonstrates that within the environment of a peer-to-peer punishment institution, in the case of Georgia, antisocial punishment emerges that undermines cooperation. Indeed, one problematic aspect of the peer punishment mechanism is that certain players may misuse the power of sanctioning incentives and thereby undermine cooperation. For instance, several PGG experiments with peer punishment documented the existence of "antisocial" punishment, whereby sanctions are extensively used against cooperators rather than free riders (Herrmann et al., (2008), Nikiforakis (2008)). Thus, while the punishment of free riders is triggered largely by violations of fairness norms, what may explain such antisocial punishment? In our experimental setup, the subjects have complete freedom – without any coordination mechanism – to determine their distribution of sanctions and to engage in antisocial punishment. Therefore, no formal institution-based collective choice exists to oversee or guide cooperation. The Herrmann et al. (2008) study equally underscores a confirmative finding, namely that the severity of antisocial punishment in a society is linked to the Rule of Law in that society. The negative correlation between antisocial punishment and quality in the Rule of Law suggests that a high-quality law enforcement system (which can be interpreted as a high degree of institutionalized cooperation) also limits antisocial punishment. Good institutions effectively lead to the solid self-governance of people who in turn manage to cooperate and who punish those who free ride.

Another plausible and commonly observed reason could be revenge, as identified by Herrmann et al. (2008). Although our experiment is not explicitly designed to measure the norm of cooperation, our evidence for the existence of considerable antisocial punishment in the PGG still suggests that no clear normative standard of cooperation and punishment seems to be at work. Thus, conclusions regarding the effectiveness of peer punishment may be misleading if they are based on institutional settings that rule out a suitable normative consensus and collective choice-based institutions that provide a formal institutional framework.

Although our experiment does not allow to assign this negative effect explicitly to peer-to-peer punishment institution, it seems to suggest that prior experience of playing PGG with punishment negatively effect on prosocial behavior in subsequent dictator game. Our experimental evidence consequently does not provide support for spillover-based theories. The lack of a positive spillover effect suggests that the peer punishment mechanism might not influencing prosociality via a change in perceived social norms. If exposure to PGG with peer-to-peer punishment impacted prosociality by changing people's individual understanding of appropriate behavior (i.e., their individual normative standard), this would also have led to changes in prosocial behavior. Therefore, our evidences call for further investigation and more cautious design of the mechanisms that could influence social preferences aimed at promoting prosociality.

Our findings are as following. Firstly, average offers of dictators are fairly similar to the typical results found in dictator game experiments. Secondly, dictators act considerably more selfishly relative to their own declared individual normative standards of fairness. Moreover, dictators become more selfish after they have had the experience of playing a public goods game with peer punishment. Thirdly, our experiment reveals a considerable share of antisocial punishment in public goods game that in turn erodes effectiveness of the peer-to-peer punishment mechanism in preventing the decay of cooperation, however, in an environment that rules out a suitable normative consensus and collective choice among the participants.

6. Conclusion

Our results suggest that the peer-to-peer punishment mechanism was ultimately an inefficient tool for enhancing cooperation in the public goods game. Peer punishment had a relatively weak disciplinary effect on free riders, in the sense of steering low contributors toward higher contributions. Moreover, we perceive substantial evidence for antisocial punishment, indicating that no common behavioral standard for individual contributions seems to emerge. However, conclusions regarding the effectiveness of the peer punishment mechanism could be misleading in an environment with institutional settings, those that rule out a suitable normative consensus and collective choice.

Our results within the dictator game do not support self-interest prediction, with many people acting fairly and generously. Dictators' average offers in our study are quite close to the typical results found in other dictator game studies. We find that dictators act significantly more selfishly relative to their own declared individual normative standard of fairness. Thus, indicating that intrinsic motives for compliance with an individual normative standard of fairness are not sufficiently strong. Next, players in the dictator game become more selfish after they have had the experience of playing a public goods game with peer punishment. Consequently, our experimental results seem to be inconsistent with spillover theory predictions. Overall, we believe that our current experimental study of social preferences and individual normative standards of fairness in a dynamic context sheds some light on the underpinnings of prosocial behavior and on the role of institutions.

References

1. Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics* 21(1): 99-131.
2. Bear, A. & Rand, D. G. (2016). Intuition, deliberation, and the evolution of cooperation. *Proceedings of the National Academy of Sciences* 113(4): 936-941.
3. Brosig, J., Riechmann, T., & Weimann, J. (2007). Selfish in the end? An investigation of consistency and stability of individual behavior. *MPRA,* Paper 2035.
4. Brosig, J, Riechmann T, Weimann J. (2017) The dynamics of behavior in modified dictator games. PLoS ONE 12(4): e0176199. https://doi.org/10.1371/journal.pone.0176199
5. Camerer, C. F. (1997). Progress in behavioral game theory. *Journal of Economic Perspectives*, 11(4), 167-188.
6. Camerer, C. F., & Fehr, E. (2004). Measuring social norms and preferences using experimental games: A guide for social scientists. In *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*, 55-95 (Ed. Henrich, J. P., Boyd, R., Bowles, S., Fehr, E., Camerer, C., & Gintis, H.). Oxford: Oxford University Press.
7. Engel, C. (2011). Dictator games: A meta-study. *Experimental Economics* 14(4): 583-610.
8. Engl, F., Riedl, A., & Weber, R. (2021). Spillover effects of institutions on cooperative behavior, preferences, and beliefs. American Economic Journal: Microeconomics, 13(4), 261-99.
9. Fehr, E. & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review* 90(4): 980-994.
10. Fehr, E. & Gächter, S. (2002). Altruistic punishment in humans. *Nature* 415(6868): 137-140.
11. Fehr, E. & Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour* 2(7): 458-468.
12. Fehr, E. & Williams, T. (2017). Creating an efficient culture of cooperation. Working paper, No 267, *University of Zurich, Department of Economics, Zurich*
13. Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and economic behavior* 6(3): 347-369.
14. Gächter, S., Nosenzo, D., Renner, E., & Sefton, M. (2010). Sequential vs. simultaneous contributions to public goods: Experimental evidence. *Journal of Public Economics* 94(7-8): 515-522.
15. Herrmann, B., Thöni, C., & Gachter, S. (2008). Antisocial Punishment Across Societies. *Science* 319: 1362-1367.
16. Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business*: S285-S300.
17. Krupka, E. L. & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association* 11(3): 495-524.
18. Mekvabishvili, R. (2021). Can Formal Institutions Lead to the Spillover Effect of Cooperation? *Theoretical Economics Letters* 11: 186-193.
19. Mekvabishvili Rati, Mekvabishvili Elguja, Natsvaladze Marine, Sirbiladze Rusudan, Mzhavanadze Giorgi & Diesadze Salome (2022). Prosocial Behavior and the Individual Normative Standard of Fairness within a Dynamic Context: Experimental Evidence. *Zenodo data repository,* https://doi.org/10.5281/zenodo.7476668
20. Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics* 92(1-2): 91-112.

21. Peysakhovich, A. & Rand, D. G. (2016). Habits of virtue: Creating norms of cooperation and defection in the laboratory. *Management Science* 62(3): 631-647.
22. Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature communications* 5: 3677.
23. Sass, M., Timme, F., & Weimann, J. (2015c). The dynamics of dictator behavior. Working Paper, No. 5348, *Center for Economic Studies and ifo Institute (CESifo).*
24. Sass, M. & Weimann, J. (2015a). Moral self-licensing and the direct touch effect. Working Paper, No. 5174, *Center for Economic Studies and ifo Institute (CESifo).*
25. Sass, M. & Weimann, J. (2015b). The peculiar power of pairs. Working Paper, No. 5246, *Center for Economic Studies and ifo Institute (CESifo).*
26. Stagnaro, M., N., Arechar, A., A., & Rand, D., G. (2017). From Good Institutions to Generous Citizens: Top-Down Incentives to Cooperate Promote Subsequent Prosociality But Not Norm Enforcement. *Cognition* 167: 212-254.