HEIDELBERG UNIVERSITY
DEPARTMENT OF ECONOMICS

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Following Social Norms, Signaling, and Cooperation in the Public Goods Game

Chi Cui

Ming Dai

Christiane Schwieren

AWI DISCUSSION PAPER SERIES NO. 746
April 2024

Following Social Norms, Signaling, and Cooperation in the Repeated Public Goods Game*

Chi Cui[1,3], Ming Dai[2], and Christiane Schwieren[2]

[1] *School of Economics and Management, Northeast Normal University, China.*
[2] *Alfred Weber Institute of Economics (AWI), Heidelberg University, Germany.*
[3] *The Key Laboratory for Applied Statistics of the MOE, Northeast Normal University, China.*

March 30, 2024

Abstract

In this paper, we experimentally investigate how sending a signal of following social norms impacts people's cooperative behavior in a repeated public goods game, where we disentangle the effect of strategy and internalization of social norms on cooperation. We find that under the signaling mechanism, less cooperative players disguise themselves in the rule-following game, but this does not decrease cooperation overall. More importantly, the signaling mechanism has a heterogeneous effect on cooperation in rule-following and rule-breaking groups: It increases cooperation in rule-following groups but decreases cooperation in rule-breaking groups. Finally, the signaling mechanism tends to offset the decline of contributions among participants in rule-breaking groups rather than rule-following groups. Overall, this paper provides a feasible way to improve social cooperation and enriches the literature on cooperation in the public goods game.

Keywords: public goods game; cooperation; signaling; internalization of social norms
JEL-Classification: C40; C92; H41

*Chi Cui is an associate professor in economics at the School of Economics and Management, Northeast Normal University, E-mail: cuic354@nenu.edu.cn; Ming Dai is a PhD candidate in economics at the Alfred-Weber Institute of Economics, University of Heidelberg, E-mail: ming.dai@awi.uni-heidelberg.de, Corresponding Address: Bergheimer Str. 58, 69115 Heidelberg, Room 00.012; Christiane Schwieren is a full professor in economics at the Alfred-Weber Institute of Economics, University of Heidelberg, E-mail: christiane.schwieren@awi.uni-heidelberg.de. The experiment was conducted in the Laboratory of Economics and Management and the Key Laboratory for Applied Statistics of MOE, Northeast Normal University. We thank Yan Chen, Te Bao, Jie Zheng, Fangwen Lu, Alexander Vostroknutov, Erik Kimbrough, Roberto Weber, Erin Krupka, and Daniel Zizzo for helpful support and comments. We also thank former and present staff, Zhiwen Lu, Keke Sun, Lulu An, Lihui Zhang, Yuanqi Tao, Jiaming Hu, Fengyi Sun for assistance.

1 Introduction

One of the most important and perennial issues for economists, sociologists, and psychologists is how to improve social cooperation. To investigate how to promote and maintain social cooperation, experimental scientists generally employ a public goods game. In these studies, much evidence shows that third-party punishment, communication, or reputation can significantly promote social cooperation (Dawes et al., 1977; Yamagishi, 1986; Isaac and Walker, 1988a; Fehr and Gächter, 2000; Masclet et al., 2003).

Many researchers have analyzed the internal and external factors motivating people to participate in voluntary cooperation, and they have found that cooperation is not only impacted by social preferences, but it is also driven by social norms (López-Pérez, 2008; Krupka and Weber, 2009, 2013; Kessler and Leider, 2012; Cappelen et al., 2013). Social norms emphasize that when people judge behavior, they compare it to an external, socially defined normative standard, and individuals internalize this process. After norms are internalized, deviations from norms, as modeled with a simple utility function, generate a utility cost (Kimbrough and Vostroknutov, 2016). When the norm is prosocial, people who suffer more from violating norms will behave more prosocially. In the public goods game, people who internalize social norms have a high willingness to cooperate. For example, Kimbrough and Vostroknutov (2016) assigned participants into groups of four based on the descending ranking of a rule-following task, and they found that cooperation in rule-following groups was significantly higher than that of rule-breaking groups, and there was no significant decline of contributions over periods in rule-following groups. Comprehensively, people who more deeply internalize social norms are more cooperative level than others.[1] In their research, K&V demonstrate that the social norm in the public goods game is the norm of conditional cooperation, so that participants will cooperate (defect) only if others cooperate (defect). But, is this always the case?

The K&V model for establishing cooperation may be more successful when a signaling mechanism is applied. Eric Posner (2009), in his book Law and Social Norms, points out that successful cooperation requires people to send a signal to others showing that they have a low discount rate and a high quality with respect to cooperation. In his point of view, following social norms can serve as a signal for improving cooperation. Based on this idea, if an individual consciously sends a signal to others by following social norms, it subsequently promotes social cooperation by activating the trust of other people. This implicates that the signaling mechanism is different from the internalization of social norms. On the one hand, people who have a low quality of cooperation (low-quality signalers) may disguise themselves by sending a high-quality signal to others although it is costly. For example, in a community, low-quality signalers can maximize their profits by matching themselves with real high-quality players who are cooperative in collective production. On the other hand, the internalization of social norms means that social norms socialize people, and they believe that following social norms is beneficial to the whole community.

In this paper, we report on an experiment where participants could choose to follow or break a social norm relevant to environmental protection in a rule-following game, and then

[1] Kimbrough and Vostroknutov (2016) grouped participants into rule-following groups and rule-breaking groups based on the ranking of rule-following task, but the process of grouping was not told to participants. Without any information of grouping, participants in rule-following groups behave intrinsic norm-following, indicating that rule-followers had a higher level of internalization of norms.

players were assigned into groups of four in the repeated public goods game, based on the descending ranking of the rule-following game. Following social norms is a tool to show signalers' personalities or qualities, but it includes two effects, namely the internalization of social norms and the strategy motivation. To disentangle two effects, we designed two treatments, where one treatment contained a signaling mechanism and the other one did not. Under the signaling mechanism, when players realize that they can benefit more from the public goods game by following social norms in the rule-following task, they will use a strategy of choosing to be more rule-following. In contrast, without such information, players who choose to follow social norms are generally intrinsic rule-followers. By comparing two situations, we can disentangle the difference between the effect of internalization and the strategy effect.

At present, most researchers have investigated social cooperation from the perspective of social preferences and norms of cooperation. This paper studies whether signaling to follow social norms can promote cooperation. The paper is organized as follows. Section 2 reviews experimental economics literature on cooperation in the public goods game. Section 3 presents the experimental design and procedures. Section 4 presents a theoretical model. Section 5 presents the analysis and main results. Section 6 concludes.

2 Literature Review

How to improve cooperation is considered an important issue for economists. Researchers often use the public goods game to study cooperative behavior (Isaac and Walker, 1988b; Santos et al., 2008; Szolnoki and Perc, 2010; Brekke et al., 2011). According to the standard rational model, being a free rider is attractive to rational individuals in the public goods game. However, a large amount of experimental evidence shows that many people are willing to sacrifice a part of their profits to achieve cooperation (Andreoni, 1988, 1995; Keser and Winden, 2000; Fishbacher et al., 2001; Haurert et al., 2002; Rege and Telle, 2004).

Yet, in repeated public goods games (Normann and Wallance, 2004), it is difficult to sustain cooperation at a high level because players intend to be free riders in later periods, which leads to significant cooperation declines (Ockenfels, 1999; Keser and Winden, 2000; Fehr and Gächter, 2001). Therefore, economists have tried different methods to deal with the problem. Researchers believe that an individual's willingness to cooperate is mainly motivated by social preferences (like reciprocity), and they have examined some mechanisms to promote cooperation, like altruistic punishment, communication, and reputation (Fehr and Gächter, 2000, 2002; Bochet et al., 2006; Carpenter, 2007; Sefton et al., 2007; Beersma and Kleef, 2011; Feinberg et al., 2014; Wu et al., 2016).

In recent work, researchers have found that people's cooperation is also motivated by social norms, indicating that people's behaviors are consistent with social norms (Ferh et al., 2002; Krupka and Weber, 2009, 2013; Kessler and Leider, 2012; Kimbrough et al., 2014). Fehr and Fischbacher (2004) reviewed evidence to suggest that human cooperation is largely based on a norm of conditional cooperation. This norm prescribes cooperation if the other group members also cooperate, whereas individuals will defect if they see others defecting. Krupka and Weber (2009) studied the impact of social norms on people's prosocial behaviors. They found that social norms strengthen people's prosocial behaviors. They compared informational effects by focusing on norms regarding prosocial behavior,[2] but they did not explore the mechanism further. Kimbrough and Vostroknutov (2016) investigated how social norms relate to people's prosocial behaviors in the public goods game; specifically, they measured the propensity of participants' compliance with social norms in a traffic-light task.[3] Then participants were grouped into rule-following groups and norm-breaking groups by the descending ranking of following traffic rules in the traffic-light task. After being grouped, participants played a repeated public goods game with fixed partners. Their results show that the cooperation of rule-followers was significantly higher than that of rule-breakers, and this cooperation sustained instead of declined; yet, this could have occurred because of unclear information of the number of rounds in the public goods game, which could have impacted players' contributions.

Signaling may be a feasible solution to increase cooperation in practice. Spence (1974) first proposed the signaling model, in which the signaler sends a signal of his/her "type" information to others by taking certain actions. Based on the signaling model, people can signal others to be more cooperative by behaving more prosocially (Zahavi, 1977; Roberts, 1998; Wright, 1999; Smith and Bird, 2000; Ginits et al., 2001; Soler, 2012; Bird and Power, 2015). Further, costly signaling theory[4] proposes that individuals often engage in behaviors that are very costly as a way of signaling honesty about themselves. Among the kinds of behaviors that can be seen as costly signaling are public philanthropy (Smith and Bird, 2000; Bereczkei et al., 2010; Haley and Fessler, 2005), risk-taking and heroism (Wilson and Daly, 1985; Hawkes, 1991; Smith, 2004), conspicuous consumption (Saad, 2007), and religious commitment. For example, Smith and Bird (2000) measured how the signal of generosity influences collective actions among Australian turtle-hunting villagers. They found that villagers' altruistic behavior serves as a signal telling other villagers that someone is a good partner and wants to continue to work together. In addition, Soler (2012) studied the impact of high-cost signals on cooperation in the public goods game. He calculated the degree of religious rituals and beliefs as a high-cost signal, where signals included attendance, giving, and participation in Candomble. (Candomble adherents perform numerous activities in a religious context that involve costs but, nevertheless, benefit the entire group, such as donating money or staying after a feast to clean the territories). Soler found that people with high religious beliefs perform more generously and are more cooperative in the public goods game than those with low religious beliefs. In another study, Hawkes (1991) developed the "show-off hypothesis" to explain the well-replicated finding that men in hunter-gather societies who are predisposed toward more risky hunting strategies end up with greater sexual access to women (Hill and Hurtado, 1996; Smith, 2004). It would be in the interest of any individual in groups composed of several families to have members of the neighboring families choose the high-risk strategy.

[2] Krupa and Weber (2009) designed an experiment including an information treatment and a focusing treatment. In the information treatment, prior to playing the game each participant observed the choices made by four previous participants, while in the focusing treatments, participants need to think about the behavior of others focuses a participant on the norm of behaving pro-socially (descriptive focus), or they will be asked what they thought other people said one should do in such a decision context (injunctive focus).

[3] Each participant makes five decisions concerning how long they wait at a sequence of red traffic lights, each of which will turn green five seconds after their arrivals. The traffic-light task creates a situation, familiar to most participants, which they are asked to follow a rule at some cost to themselves.

[4] Costly signaling theory proposes that animals (including humans) may send honest signals about desirable personal characteristics and access to resources through costly biological displays, altruism, or other behaviors that would be hard to fake (McAndrew, 2021).

While neighbors who are provisioners work only for their spouses and children, neighbors who gamble for occasional bonanzas work for everyone in the community and not just for their own households. Even though a woman might prefer to be a provisioner and to have a provisioning husband, she would prefer others in the community to bring in jackpots, that is, to behave as show-offs (Hawkes, 1991).

The success of the costly signaling theory depends on the cost of signaling. The "cost" of a signal to the sender is a reliable way of confirming whether the signal is honest, so costly signaling is very much about truth in advertising. A low-quality signaler who attempts to fake a high-quality signal will deplete whatever resources that he may have available, leaving the signaler in such a vulnerable position that the strategy will prove to be counterproductive. Conversely, a high-quality signaler has resources to burn and can easily afford a high-quality signal, so the adaptive benefits will outweigh the costs (Grafen, 1990). The costs of these signals are relatively high, and they are not affordable for everyone who would like to cooperate, but if the cost of a signal is low enough, low-quality signalers probably send fake signals of high quality to receivers. What would happen to cooperation if the costless signal induced a disguise? If the costless signal can significantly improve cooperation, it is valuable even though it induces a disguise.

Following social norms may serve as a costless and feasible signal. Norms can signal intentions, aspects of personal character, or membership in a group. Although the behaviors themselves are of little consequence, they have important reputational implications. Dress codes are often used to signal membership in specific groups or the holding of particular preferences, such as veiling by Muslim women (Carvalho, 2013), "hanky codes" among gay men, and tattoos on criminals (Gambetta, 2009). Posner (2009) pointed out in the book Law and Social Norms that the establishment of cooperation requires people to send a signal to each other to show that they have a low discount rate and are worthy of cooperating with others. In Posner's view, following social norms is a signal that reveals the good qualities of signalers, so it is easier for receivers to identify good people and start cooperation. Other signals connote general rather than specific traits: Observing fine points of etiquette, showing up on time, speaking in turn, and displaying appropriate degrees of deference are often taken to be signs of reliability and trustworthiness (Posner, 2009). Therefore, we investigate whether people can make decisions for cooperating with others by sending signals of following social norms that are relevant to environmental protection and further examine whether a disguise takes place before cooperation starts.

This paper modifies the experimental design of Kimbrough and Vostroknutov (2016, 2018). To study the impact of signaling with social norms on cooperation, we analyze the difference between the signaling mechanism and internalization of social norms. The main contributions of this paper to the literature are as follows: First, this paper employs a really simple and costless social norm that is relevant to environmental protection, as a signal to improve cooperation. By comparing the internalization of social norms, we find the signaling mechanism is a good way to improve cooperation by matching participants with those who have a similar level of following social norms. Thus, this work further enriches the research of Krupka and Weber (2009) and Kimbrough and Vostroknutov (2016, 2018). Second, we elicit the effect of the internalization of social norms and the strategy motivation of signaling by comparing the signaling treatment group with other treatments groups that did not use the signaling mechanism. More importantly, we also find the heterogeneous effect of the signaling on different groups, which offers us a perspective of looking at the interests and complexity of the signaling effects. Third, to examine whether disguising actions under the signaling mechanism takes place, we employed a dice game for measuring the extent to which participants cheat; this allowed us to distinguish differences between the signaling mechanism and the internalization of social norms. Honest individuals will follow social norms, while dishonest individuals might follow social norms for pretending to be good people. In previous studies, researchers could not distinguish between participants who were intrinsic rule-followers and those who were rule-breakers, but here we identify them based on what they did in the dice game before playing the rule-following game. Finally, instead of assessing one-shot cooperation with a signaling mechanism as has been done in previous literature, in this paper here, we observe the effect of continuous cooperation in a repeated public goods game involving signaling.

3 Experimental Design and Procedure

We designed an experiment that contains a series of games to investigate how following social norms as a signal impacts people's cooperation in the repeated public goods game and to disentangle effects of the internalization of social norms and the strategy motivation. We now present the economic environments and our experimental procedure.

3.1 Experimental Design

The experiment was conducted at the Economic Management Laboratory and the Key Laboratory for Applied Statistics of MOE, Northeast Normal University, in the winter semester of 2018, the summer semester of 2019, and the winter semester of 2021. The experiment we conducted over this long time frame because of the COVID-19 pandemic in December 2019. We recruited 272 participants from Northeast Normal University, and they were randomly assigned into each treatment. Participants' academic backgrounds included social sciences, humanities, and engineering. The number of participants in a session was 16, and one session took on average one hour. Average earnings were 20.5 RMB (3.1 US dollars) plus a show-up fee of 15 RMB (2.23 US dollars).

Based on the designs of Kimbrough and Vostroknutov (2016, 2018), we introduced a signaling mechanism to the experiment. The experiment consisted of four parts: the first one was a dice game for measuring the participants' degree of honesty, the second one was a rule-following game, the third one was a standard linear public goods game (Leyard, 1994), and the last one was a short questionnaire. Participants first participated in the dice game,5 and then they played the rule-following game. Finally, they played the public goods game[5] for ten rounds with fixed partners. At the end of the experiment, they filled in a questionnaire including demographics. Participants' final profit was the sum of income from the three games. The experiment was programmed in z-Tree (Fischbacher, 2007), and the entire experimental process was completed on computers.

The dice game was used to measure whether or not a group of participants cheated. Generally, there is a correlation between the degree of honesty of a group and prosociality. Honest people are more likely to behave more prosocially, yet this does not mean that someone who is prosocial will behave honestly in the dice game. The dice game was based on an

[5] We put the dice game before the rule-following game and the repeated public goods game in order to detect the pure degree of honesty that is not impacted by other games.

experimental design by Fischbacher and Föllmi-Heusi (2013) and Gächter and Schulz (2016). In this game, participants rolled a six-sided die in an opaque cup twice, and they had to input the number of the first roll and the corresponding payoff to the computer. Dice rolls could not be seen by anyone except the participant himself/herself. Participants were paid according to the number they reported. Here, the payoff would equaled 5, 10, 15, 20, or 25 tokens if the dice number that came up corresponded to the payoff amount, and it equaled 0 token if the dice number that came up was a 6 (See Table 1). Many people desire to maintain an honest self-image, and lying about a dice roll jeopardizes this self-image; but, bending rules might not. Bending the rules might be done by reporting the higher of two rolls, rather than the first roll, as instructed.[6] We modified the dice game of Fischbacher and Föllmi-Heusi (2013) and Gächter and Schulz (2016) in this way: To balance incomes among the three games, we resized the payoff of the dice game to be four times bigger than that in the original design (Gächter and Schulz, 2016). Participants were instructed to input the first rolled number and the corresponding payoff into the computer. The income in this game was the payoff participants input multiplied by 0.1.

Table 1: Numbers and payment in the dice game

Number	1	2	3	4	5	6
Payoff	5	10	15	20	25	0

Next, the rule-following game (the RF game) was used to measure the participants' propensity to follow a norm related to environmental protection. According to the experimental design of Kimbrough and Vostroknutov (2016), we directly elicited the degree of following a social norm through a rule-following game. In this game, as depicted in Figure 1 each participant was able to put rubbish balls in the trash can (Box A) or throw rubbish balls on the ground (Box B). With a total of 100 rubbish balls, participants decided to allocate some number of rubbish balls to the trash can and throw the other rubbish balls on the ground.

Participants were told the social norm that *"People should put all rubbish (balls) into the trash can (Box A)."* This is the simplest social norm relevant to environmental protection for participants, and protecting the environment is one of the most important policies in China. Since 2013, China's government has released a series of regulations and laws for environmental protection, including air pollution, water pollution, and garbage classification, and, meanwhile, the government invested more money in recycling industries. As for Chinese students, environmental protection is common knowledge in daily life, so putting rubbish balls in the trash can is easy to understand and a common norm. The price of a rubbish ball is different between Box A and Box B. The price is 0.25 tokens per ball in Box A and 0.4 tokens per ball in Box B. This means that participants are rule-followers when all 100 balls (rubbish) are put into Box A (trash can). If 100 balls are put into Box B, this shows that this person is a complete rule-breaker. The total profit of this game is the sum of profits in Box A and

[6] In the research of Gächter and Schulz (2016), reporting the better of two rolls implies the "justified dishonesty" benchmark: Claims of 6 should occur in $\frac{1}{36} \approx 2.8\%$ of the cases (after rolling (6, 6)); claims of 1 should occur in $\frac{3}{36} \approx 8.3\%$ (after (6, 1) or (1, 6) or (1, 1)); claims of 2, 3, 4 and 5 should occur in 13.9%, 19.4%, 25% and 30.6% of cases, respectively.

Box B.

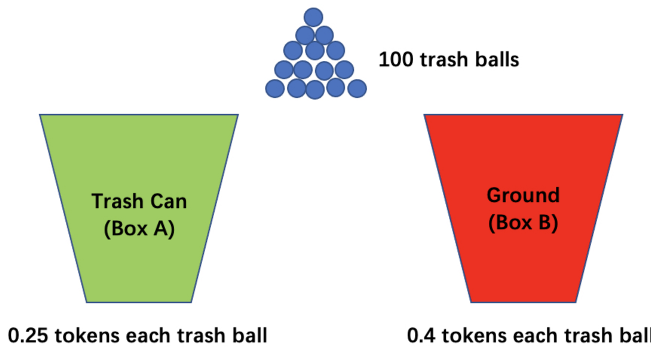
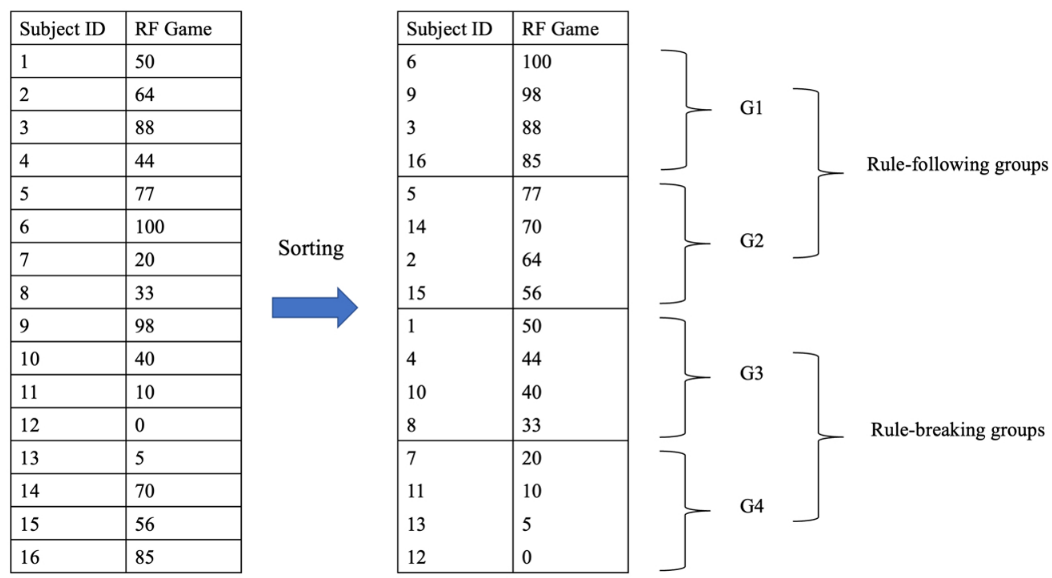


Figure 1: Rule-following game

Finally, the repeated public goods game (PGG) was used to measure people's cooperative behavior. Each participant was given an endowment of 20 tokens in each round, and participants were randomly assigned into groups of four. Participants played the PGG for ten rounds, and the composition of each four-participant group did not change throughout the entire game. Each participant was given a private account and a public account, and they were allowed to allocate their endowment between the two accounts in each round. If one participant chose to invest X tokens into the public account and the other teammates chose to invest Y, A, and B tokens, respectively, then the participants' income would be $20 - X + 0.4 \cdot (X + Y + A + B)$ in this round. The total income earned by the participants in this experiment was the sum of incomes from ten rounds multiplied by 0.1.

3.2 Experimental Procedure

The experiment had three treatments, namely the baseline treatment (BT), internalization of norms treatment (INT), and signaling treatment (ST). All participants were randomly assigned to each treatment. The experiment had 17 sessions, and each session consisted of 16 participants. All participants were randomly assigned into BT, INT, and ST, and they were assigned into four groups (G1, G2, G3, and G4) of four teammates in the repeated PGG. Participants in the in BT were randomly assigned into four subgroups of four teammates in the PGG. But participants in the INT or ST were assigned into subgroups of four by the descending ranking of following social norms in the RF game. Specifically, participants who put more rubbish balls in the trash can were more likely to join groups where rule-followers were dominant. In INT and ST, participants in G1 and G2 had a higher propensity to follow norms than those in G3 and G4, so we considered G1 and G2 to be rule-following groups, while G3 and G4 were considered rule-breaking groups. Figure 2 shows an example of how participants were grouped according to their behavior in the RF game.



Subject ID	RF Game
1	50
2	64
3	88
4	44
5	77
6	100
7	20
8	33
9	98
10	40
11	10
12	0
13	5
14	70
15	56
16	85

Subject ID	RF Game
6	100
9	98
3	88
16	85
5	77
14	70
2	64
15	56
1	50
4	44
10	40
8	33
7	20
11	10
13	5
12	0

Figure 2: An example of grouping rule in INT and ST

Baseline Treatment (BT): Before the dice game started, experimenters told participants that they would be participating in a game of rolling a dice using an opaque cup and they were taught how to play it. After the dice game, participants took part in an RF game where they were told to allocate 100 rubbish balls between Box A (trash can) and Box B (on the ground). When the RF game ended, experimenters told participants to play the PGG for ten rounds with fixed teammates. Participants start to play the PGG after being randomly assigned into groups of four (see Figure 3).

Internalization of Norms Treatment (INT): The dice game and the RF game in the INT were the same as in the baseline treatment (BT). The grouping principle in the repeated PGG was similar to the design of Kimbrough and Vostroknutov's (2016, 2018). After finishing the RF game, participants were assigned into groups of four by the descending ranking of following social norms in the RF game (based on the number of balls put in Box A), but they were not told how they were grouped. After grouping, experimenters instructed the participants on how to play the public goods game, which participants then played for ten rounds with fixed partners.

Signaling Treatment (ST): The dice game in ST was the same as that of BT and INT. When participants finished the dice game, experimenters instructed participants to first participate in an RF game and then play the PGG for ten rounds. Experimenters told participants that they would be grouped into groups of four in the PGG based on their decision in the previous game (the RF game). The grouping principle was stated as follows: *"You will be matched with the three partners who have the same type as you, in other words, the number of balls they put into Box A is close to yours."* Therefore, each participant was able to send a signal to others by following a rule or breaking a rule, indicating that he/she is a specific kind of person (a potentially cooperative person or free rider). The grouping principle in this treatment was the same as in the INT. After being grouped, participants played the PGG.

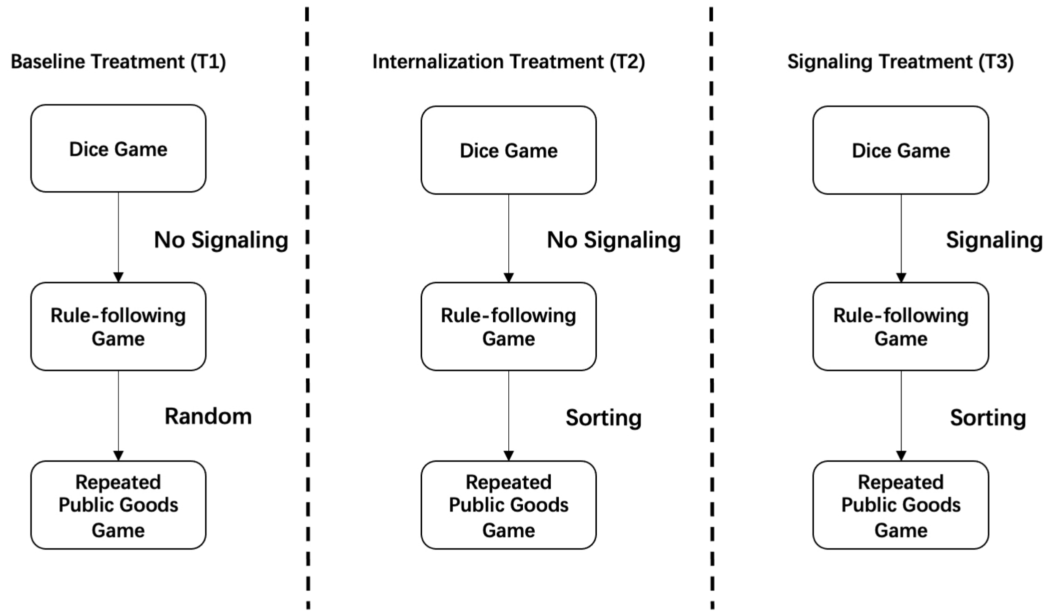


Figure 3: Treatments and experimental procedure

4 Theoretical Model

4.1 Norm-Dependent Utility

The norm-dependent utility model reflects the intuition behind the interpretation of the observed relationship between behavior in the rule-following game and in the public goods game (Kimbrough and Vostroknutov, 2016). We assume subject i in the RF game gets linear consumption utility from money and incurs costs from not following the norm (Kimbrough and Vostroknutov, 2016, 2018). The costs are higher the fewer the number of balls in the trash can (Box A), meaning subject i breaks the rule. So, the utility from throwing x rubbish balls on the ground (Box B) can be written as follows:

$$U_i(x) = 0.4x + 0.25 \cdot (100 - x) - \phi_{RF,i} \cdot g(x)$$
$$\Rightarrow U_i(x) = 25 + 0.15x - \phi_{RF,i} \cdot g(x)$$

Here, $\phi_{RF,i} > 0$ is the propensity of i to follow norms, and $g : \mathbb{R}_+ \rightarrow [-1, 1]$ is a function that assigns a normative social appropriateness (inappropriateness) to each action x which represents the number of rubbish balls in Box B (breaking the rule). In our experiment, subject i has 100 rubbish balls in the RF game, so $0 \leq x \leq 100$. The price per ball in the trash can (Box A) is 0.25 tokens, but the price per ball thrown on the ground (Box B) is 0.4 tokens. Function $g(x)$ is assumed to capture the unique norm shared by all members of the society, which is assumed to be independent of the individual parameter $\phi_{RF,i}$. The profit in the rule-following game is the sum of the profit in Box A and the profit in Box B $(0.4x+0.25(100-x))$. Meanwhile, x in the Box B determines subjects to suffer from disutility or gain positive utility.

In the RF game, $-\phi_{RF,i} \cdot g(x)$ is decreasing in x, so following the rule can produce positive utility to subjects, and disutility can come from breaking the rule. For example, if subject i puts 100 rubbish balls in Box A, indicating he or she is completely following the rule of environmental protection, we will have a positive utility $\phi_{RF,i}$ from following the rule, but there is a disutility $-\phi_{RF,i}$ when subject i puts 100 rubbish balls in the Box B (completely breaks the rule). However, the profit from breaking the rule $(25 + 0.15x)$ is linear and increasing but $-\phi_{RF,i} \cdot g(x)$ is decreasing in x, so the total utility $U_i(x)$ is not a linear function but a sum of $25 + 0.15x$ and $-\phi_{RF,i} \cdot g(x)$. Thus, subject i should make a tradeoff between following the rule and breaking it. Suppose that subject i maximizes his or her utility. Then, given the appropriate assumption on the shape of $g(x)$, we would have the following equation $\phi_{RF,i} = 0.15/g'(x_1^*)$, where x_1^* is the optimal choice (observed by us in the RF game). Thus, we obtain a positive monotonic relationship between unobserved $\phi_{RF,i}$ and observed x_1^*. In this sense, we consider the observations in the RF game as a proxy for $\phi_{RF,i}$. We can say that $g(x)$ is dependent on the internalization of social norms. If individuals have a high-level internalization of social norms, they will follow social norms without any hesitation although it will decrease their profits.

Hypothesis 1: Intrinsic rule-followers follow social norms in the rule-following game to a greater degree than intrinsic rule-breakers, which means that intrinsic rule-followers will put more rubbish balls in the trash can than intrinsic rule-breakers.

4.2 Norm-Dependent Utility Under the Signaling Mechanism

Intrinsic rule-followers (high quality) have a high propensity to follow social norms, while rule-breakers (low quality) have a low propensity, so we assume that $\phi_{RF}^H > \phi_{RF}^L > 0$. Subject i has an idea about his or her own propensity to follow social norms but cannot be observed by other subjects, but subjects can observe the number of rubbish balls thrown on the ground, which is determined by one's propensity to follow social norms. We assume that rule-followers will have x_1 rubbish balls on the ground ($100 - x_1$ rubbish balls in the trash can), and rule-breakers will have x_2 rubbish balls on the ground, which means that $x_1 < x_2$. We have assumptions as follows

1. Subject i is a rule-follower with probability of P_0, otherwise a rule-breaker with probability of $1 - P_0$.
2. Subjects can choose the action of following social norms with x_1 or breaking social norms with x_2.
3. Subject i will get an additional utility $f(x)$ which represents the quality of the signal. The utility from having x balls in Box B could be rewritten as follows:

$$U_i(x) = [25 + 0.15x - \phi_{RF,i} \cdot g(x)] \cdot f(x)$$

Here, $f : \mathbb{R}_+ \to [-1, 1]$ is a function that represents the quality of the signal by the signaler. The function $f(x)$ is strictly decreasing in x, indicating that subject i sends a low-quality signaling to other people when he/she puts rubbish balls in Box B. The higher quality of the signaling that subject i sends to others, the greater probabilities he/she has of meeting intrinsic rule-followers in the next game (the repeated public goods game), so that he/she can maximize the profit of conditional cooperation in the PGG. In the following, we discuss the separation equilibrium and the pooling equilibrium, where we examine whether subjects have a disguise in the rule-following game.

4.2.1 Separation Equilibrium

If there exists a separation equilibrium, there must be constraints as follows

$$\begin{cases} (25+0.15x_1-\phi_{RF}^{H}g(x_1))\cdot f(x_1) > (25+0.15x_2-\phi_{RF}^{H}g(x_2))\cdot f(x_2), \\ (25+0.15x_1-\phi_{RF}^{H}g(x_1))\cdot f(x_1) > 0, \\ (25+0.15x_2-\phi_{RF}^{L}g(x_2))\cdot f(x_2) > (25+0.15x_1-\phi_{RF}^{L}g(x_1))\cdot f(x_1), \\ (25+0.15x_2-\phi_{RF}^{L}g(x_2))\cdot f(x_2) > 0 \end{cases} \tag{1}$$

where ϕ_{RF}^{H} and ϕ_{RF}^{L}, respectively, represent the propensity to follow the rule of a high-quality player and low-quality player in the RF game. If there exists a separation equilibrium, any combination (x_1, x_2) will meet the above inequations at any time, while there does not exist such a separation equilibrium only if we find a combination which breaks the above inequations. We set (1). We set $(x_1, x_2) = (0, 100)$, which indicates that rule-followers put all the rubbish balls in the trash can (completely follow social norms) while rule-breakers put all the rubbish balls on the ground (completely break social norms). We find that

$$\begin{cases} 32.5 > \phi_{RF}^{H}, \\ \phi_{RF}^{H} > -25, \\ -40 > 25, \\ \phi_{RF}^{L} > 40 \end{cases} \tag{2}$$

From the above inequations, we know that one of the inequations is wrong ($-40 > 25$), and $\phi_{RF}^{L} > 32.5 > \phi_{RF}^{H} > -25$ is a contradiction to one of the most important assumptions $\phi_{RF}^{H} > \phi_{RF}^{L} > 0$, so it is impossible to have a separation equilibrium. This means that rule-followers and rule-breakers will not take different actions in the RF game.

4.2.2 Pooling Equilibrium

Now that there is no separation equilibrium, we discuss the pooling equilibrium. If there is a pooling equilibrium, it indicates that rule-followers and rule-breakers will take similar actions. It is profitable for low-quality players to act like high-quality ones, so we also have the below equations as follows:

$$\begin{cases} (25+0.15x_1-\phi_{RF}^{H}g(x_1))\cdot f(x_1) > (25+0.15x_2-\phi_{RF}^{H}g(x_2))\cdot f(x_2), \\ (25+0.15x_1-\phi_{RF}^{H}g(x_1))\cdot f(x_1) > 0, \\ (25+0.15x_2-\phi_{RF}^{L}g(x_1))\cdot f(x_1) > (25+0.15x_2-\phi_{RF}^{L}g(x_2))\cdot f(x_2), \\ (25+0.15x_2-\phi_{RF}^{L}g(x_2))\cdot f(x_2) > 0 \end{cases} \tag{3}$$

Low-quality players put $100 - x_1$ rubbish balls in the trash can, which is the same as the number of the high-quality players. We set $(x_1, x_2) = (0, 100)$, and we get the following results:

$$\begin{cases} 25 > -40, \\ \phi_{RF}^{H} > -25, \\ 25 > -40, \\ \phi_{RF}^{L} > 25 \end{cases} \tag{4}$$

The pooling equilibrium exists only if the above inequations (3) can be met with ϕ_{RF}^{H} and ϕ_{RF}^{L}. In other words, $\phi_{RF}^{H} > \phi_{RF}^{L} > 25$ is a sufficient and necessary condition for a pooling equilibrium where low-quality signalers take an action indicative of high-quality signalers. Low-quality players are willing to disguise their nature in the RF game in order to join rule-following groups in the PGG, while high-quality signalers will not change their strategy of being a rule-follower. Thus, low-quality signalers will put much more rubbish balls in the trash can, which signals a cooperative partner.

Hypothesis 2: Under the signaling mechanism, there does not exist a separation equilibrium in the rule-following game, but the pooling equilibrium can be found, which indicates that some subjects take an action aimed to disguise.

4.2.3 Repeated Public Goods Game

Subjects realize that following social norms contains information about their own characteristics or personalities when others receive that signal. Subject i has such a belief about other members who have a close number of rubbish balls in the trash can in the RF game, which means that he/she realizes that $\phi_{RF,i}$ is close to that of other members, and group members prefer to follow social norms or break social norms. Subject i only can conjecture the parameter $\phi_{RF,i}$ by the number of rubbish balls in the trash can. In the repeated PGG, Subject i has a belief about other members' $\phi_{RF,-i}$ in the RF game, which can be considered a prediction for the norm of conditional cooperation $\phi_{RF,-i}$ of other members; thus, we have a function as follows:

$$\phi_{PGG,i} = F(\phi_{RF,i}, N(t-1) \cdot C_{-i,t-1} | \phi_{RF,i})$$

where $\phi_{RF,-i}$ represents subject i's belief about others' propensity to follow social norms in the RF game, and $\phi_{PGG,-i}$ represents subject i's belief about others' propensity to follow the norm of conditional cooperation in the repeated PGG that subject i can conjecture based on $\phi_{RF,-i}$ and the mean lagged contributions of other members $C_{-i,t-1}$. The function $F(\cdot)$ is a Bayesian function which will be updated with the latest period mean contributions of other members, so when $t \geq 2$, $N(t-1) = 1$, but if $t = 1$, $N(t-1) = 0$, indicating that $\phi_{PGG,-i}$ only depends on $\phi_{RF,-i}$ at the first-period contribution in the repeated PGG. $F(\cdot)$ is increasing in $\phi_{RF,-i}$ and $N(t-1)C_{-i,t-1}$ conditional on $\phi_{RF,i}$. The function $F(\cdot)$ is also increasing in $\phi_{RF,-i}$, that there is a greater probability that subject i matches himself/herself with high-quality partners only if he/she has a higher level of propensity to follow social norms in the RF game.

Assume that we have an n-agent, linear, voluntary contributions to PGG with T periods. $C_{it} \in [0, 20]$ is the number of contributions to the public account by subject i in period t. The profit at period t to subject i is:

$$\pi_{i,t}(C_{i,t}, C_{-i,t}) = 20 - C_{i,t} + \alpha \sum_{j=1,\ldots,n} C_{j,t}$$

where $0 < \alpha < 1$ and $n\alpha > 1$, which indicates that the profit from full cooperation is greater than that of full defection. Each subject has an endowment of 20 tokens in each period. We refer to the norm-dependent utility of the PGG in Kimbrough and Vostroknutov (2016), so the distance between contributions is defined as the distance between corresponding real numbers. The utility of subject i in period 1 is:

$$u_{i,1}(C_{i,1}, C_{-i,1}) = \pi_i(C_{i,1}, C_{-i,1}) - \phi_{PGG,i} \cdot g(||\eta - x_{i,1}||)$$

Here, $\phi_{PGG,i}$ and $g(\cdot)$ are the same objects as in the previous section, and η represents the action considered the most appropriate by all subjects. Under the signaling mechanism, at the first period, the $Contribution_1$ of subject i will be impacted by the belief about other members' propensity to follow social norms in the RF game, so we can modify the first-period utility here:

$$u_{i,1}(C_{i,1}, C_{-i,1}) = \pi_i(C_{i,1}, C_{-i,1}) - \phi_{PGG,i} \cdot \phi_{PGG,-i,1} \cdot g(||\eta - C_{i,1}||)$$

where $\phi_{PGG,-i,1}$ is increasing in $\phi_{RF,-i,1}$,indicating that the belief of others' propensity of following the norm of conditional cooperation at the first period is only dependent on the belief about others' propensity of following the norm in the RF game. $\phi_{PGG,i}$ is the propensity of following the norm of conditional cooperation, which is considered to be constant because of the individual preference stability. Subjects in rule-following groups have a greater assessment of others' $\phi_{PGG,-i,1}$ than those in rule-breaking groups at the first period, so participants in rule-following groups will contribute more than those in rule-breaking groups.

Hypothesis 3: Under the signaling mechanism, subjects in rule-following groups contribute more to the public goods game in the first-period contributions than those in rule-breaking groups.

The socially appropriate action in any period depends on the prior play of other members. As Kimbrough and Vostroknutov (2016) did, we introduce a norm of conditional cooperation to formalize the intuition that people are not blind adherents to some action, who will continue to choose it under any circumstances. Instead, individuals choose actions conditional on past events. We can rewrite the above utility as follows:

$$u_{i,t}(C_{i,t}, C_{-i,t}) = \pi_i(C_{i,t}, C_{-i,t}) - \phi_{PGG,i} \cdot \phi_{PGG,-i,t}(\phi_{RF,-i}, C_{-i,t-1}) \cdot g(||\eta - C_{i,1}||)$$

where $u_{i,t}(C_{i,t}, C_{-i,t})$ depends on the past contributions of other members, which will determine the belief about others' propensity of following the norm of conditional cooperation ($\phi_{PGG,-i,t}$), and $\phi_{PGG,-i,t}$ is increasing in $\phi_{RF,-i}$ and $C_{-i,t-1}$ ($t \geq 2$). When the period $t \geq 2$, the belief of $\phi_{RF,-i}$ in the RF game is determined before the first-period contribution, so $\phi_{RF,-i}$ is like a constant and not changed again, but the lagged contributions of other members always change over periods. Now the belief about others' propensity of following the norm of conditional cooperation ($\phi_{PGG,-i}$) is only determined by the past event $C_{-i,t-1}$. Subjects in the rule-following groups have a greater belief about the RF game than those in the rule-breaking groups, so $\phi^{RF}_{PGG,-i,t} > \phi^{RB}_{PGG,-i,t}$, which indicates that $C^{RF}_{t-1} > C^{RB}_{t-1}$, and we can also get the result that $C^{RF}_{t-1} > C^{RB}_{t-1}$. Subjects in rule-following groups contribute more than those in rule-breaking groups over periods.

Hypothesis 4: Contributiont of rule-following groups is greater than that of rule-breaking groups at period t ($t \geq 2$).

5 Results

5.1 Description of Data

Table 2 provides information on the average rate of contributions in the PGG, the number of groups, the average number of balls in Box A (the RF game), and the average payoff for the dice game. The number of participants in the three treatments was 96 participants in the BT, 112 participants in the INT, and 112 participants in the ST.[7] The average rate of contributions in the PGG was 30.9% in BT, 35.0% in INT, and 34.4% in ST. In the RF game, on average, participants put 58 balls into Box A in BT, 54 balls in Box A in INT, and 53 balls in Box A in ST. In the dice game, the payoff in ST was the greatest (13.8), which was more than the participants in BT (14.4) and INT (15.8) earned.

Table 2: Data statistics in all treatments

Treamnts	Groups	Participants	Public goods game	Rule-following game	Dice game
BT	24	96	0.309	58.39	13.85
			(0.336)	(36.88)	(8.223)
INT	28	112	0.350	54.11	14.38
			(0.336)	(35.22)	(8.728)
ST	28	112	0.344	52.90	16.80
			(0.322)	(36.30)	(7.985)

Note: Standard error in parentheses. BT, INT, and ST represent the baseline treatment, the internalization of norms treatment, and the signaling treatment, respectively.

[7] There are less participants in Baseline Treatment (BT) than the other treatments, because the Baseline Treatment does not include the difference between rule-following and rule-breaking groups, thus required less data here than the other treatments.

5.2 Rule-Following Game and Dice game

We performed a Wilcoxon-Mann-Whitney test on the number of rubbish balls put into Box A of the RF game in the three treatments. We found that BT is not significantly different from INT (Wilcoxon-Mann-Whitney test: z=0.597, p=0.551) or ST (Wilcoxon-Mann-Whitney test: z=0.837, p=0.403). There is no significant difference between INT and ST (Wilcoxon-Mann-Whitney test: z=0.212, p=0.832).

Because the grouping principle in BT was a random assignment, participants in this condition cannot be split into rule-following groups and rule-breaking groups. As Figure 4 shows, there is a significant difference between rule-following groups and rule-breaking groups in both INT and ST, whereas there is no difference between rule-following and rule-breaking groups in BT. In the internalization of norms treatment (INT), rule-followers on average put 82 balls into Box A while rule-breakers in INT put fewer than 26 balls into Box A (Wilcoxon-Mann-Whitney test: z=8,517, p=0.000), indicating that rule-followers have a higher propensity to follow the norm of putting trash in the trash can than rule-breakers. This supports Hypothesis 1 that rule-followers have a higher propensity to follow the norm in the RF game than rule-breakers. In the signaling treatment (ST), 83 rubbish balls were put into Box A in rule-following groups, and 23 balls were put into Box A in the rule-breaking group (Wilcoxon-Mann-Whitney test: z=8.722, p=0.000). When comparing rule-following groups (rule-breaking groups) between INT and ST, we found no significant difference between INT-RF and ST-RF groups (Wilcoxon-Mann-Whitney test: z=-0.276, p=0.782) or INT-RB and ST-RB[8] groups (Wilcoxon-Mann-Whitney test: z=0.755, p=0.452).

To examine whether the signaling mechanism impacted decision-making in the RF game, we compared the RF game within subgroups (G1, G2, G3, or G4). If some participants disguised themselves under the signaling mechanism, there should be a difference in the number of rubbish balls that were put in the trash can within subgroups. Each session had 16 participants who were assigned into four groups of four (G1, G2, G3, and G4) based on the descending ranking of the RF game. Rule-following groups include G1 and G2, and rule-breaking groups include G3 and G4 (see Figure 5). Table 3 shows that the mean of rubbish balls in the trash can in rule-following groups was greater than that of rule-breaking groups, indicating that participants want to send high-quality signals by increasing rubbish balls in Box A, indicating that participants can match themselves with high-quality teammates by increasing the rubbish balls in the trash can (showing they are rule-followers), which means that the signaling mechanism induced participants to take actions of sending signals.

Result 1: The signaling mechanism indeed made participants send signals to others by increasing their degree of following the social norm.

[8] The terms INT-RF and INT-RB refer to the rule-following and rule-breaking groups, respectively, in the internalization of norms treatment. Similarly, ST-RF and ST-RB refer to the rule-following and rule-breaking groups, respectively, in the signaling treatment.

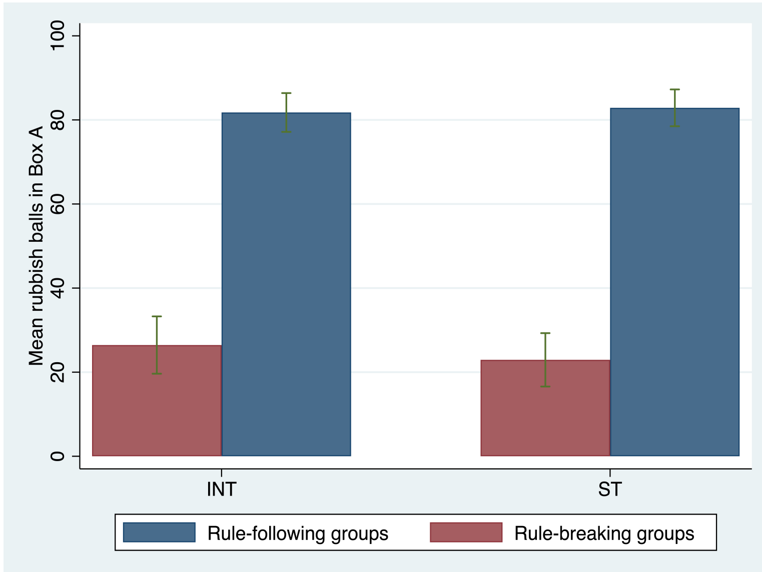


Figure 4: Means of rubbish balls in the rule-following and rule-breaking groups

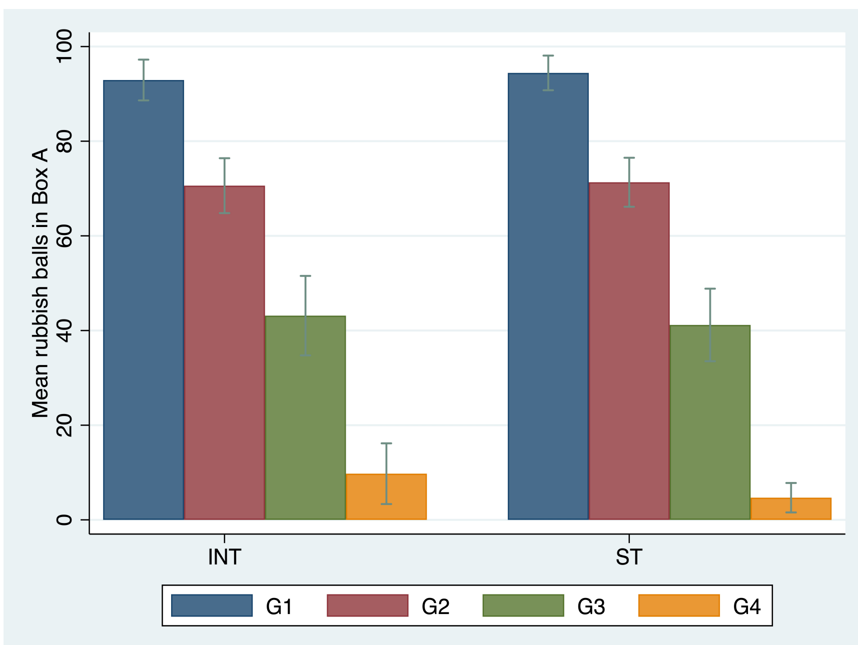


Figure 5: Means of rubbish balls in subgroups

Table 3: T-test on rubbish balls in ST

Groups	Subgroups	Obs.	INT	ST	t-value	p-value
Rule-following groups	G1	56	92.9	94.4	-0.545	0.294
	G2	56	70.6	71.3	-0.189	0.426
Rule-breaking groups	G3	56	43.1	41.2	0.355	0.362
	G4	56	9.75	4.67	1.459	0.075*

Note: Observations=2 treatments×7 groups×4 players. INT and ST, respectively, are the internalization of norms treatment and the signaling treatment. * Significance at 10%, ** significance at 5%, *** significance at 1%.

To examine whether the signaling mechanism induces people to perform a disguise action in the RF game, we employed the dice game before the RF game (Fischbacher and Föllmi-Heusi, 2013; Gächter and Schulz, 2016). We hypothesized that for the dice game in INT, rule–following groups will behave fully honestly, while rule-breaking groups will follow justified dishonesty, so there will be a significant difference in cumulative distribution between rule-following groups and rule-breaking groups. Thus, when a disguise takes place in ST, it means that some low-quality signalers can join the rule-following groups by putting more rubbish balls in the trash can, which will decrease the average degree to honesty in rule-following groups of ST, such that the cumulative distribution in rule-following groups is close to that of justified dishonesty.[9].

As Figure 6 shows, in INT, there is a significant difference in honesty between rule-following and rule-breaking groups (Wilcoxon-Mann-Whitney test: z=-1.732, p=0.084), indicating that intrinsic rule-followers behave differently than intrinsic rule-breakers. We performed a cumulative distribution test on the dice game in the rule-following and rule-breaking groups with the Kolmogorov-Smirnov test. We found that there is a significant difference in the cumulative distribution between the honesty degree of rule-following groups and justified dishonesty (Kolmogorov-Smirnov test, d=0.217, p=0.002), and the cumulative distribution of rule-following groups is not significantly different from full honesty (Kolmogorov-Smirnov test, d=0.071, p=0.592). The distribution in rule-breaking groups is significantly different from justified dishonesty (Kolmogorov-Smirnov test, d=0.191, p=0.008), and there is a significant difference between the distribution in rule-breaking groups and full honesty (Kolmogorov-Smirnov test, d=0.185, p=0.012), which means that participants in rule-breaking groups are neither fully honest nor completely dishonest. The correlation between the dice game and the RF game is significantly negative (PCC=-0.206, p=0.029), indicating that intrinsic rule-followers are more inclined to be honest than intrinsic rule-breakers (b=-0.5, p=0.096).

Contrary to INT, there is no significant difference in the distribution between rule-following and rule-breaking groups in ST (Wilcoxon-Mann-Whitney test: z=-1.224, p=0.224), which means that the degree of honesty in rule-following groups is close to that of rule-breaking

[9] Justified dishonesty is different from honesty or dishonesty, but it is still a kind of cheating action. In our research, participants are told to roll the dice twice and input the first number (and that number's corresponding payoff) in the computer. Yet, some participants may choose to put the better number and the corresponding better payoff between two dice rolls, which is called *justified dishonesty* (Gächter and Schulz, 2016).

groups. In ST, the distribution in rule-following groups is significantly different from justified dishonesty (Kolmogorov-Smirnov test, d=0.173, p=0.015); meanwhile, the distribution in rule-following groups is significantly different from full honesty (Kolmogorov-Smirnov test, d=0.161, p=0.039), indicating that the cumulative distribution is between full honesty and justified dishonesty, which is different from result of rule-following groups in INT where rule-following groups is consistent with full honesty but significantly different from justified dishonesty. It implies that more cheaters joined rule-following groups in ST than INT. On the other hand, there is a significant difference between the level of honesty in rule-breaking groups and full honesty (Kolmogorov-Smirnov test, d=0.286, p=0.000), and it is not different from justified dishonesty (Kolmogorov-Smirnov test, d=0.083, p=0.413). These results illustrate that disguises indeed happen in ST, especially in the rule-following groups, which is consistent with Hypothesis 2.

Result 2: Intrinsic rule-followers were consistent with full honesty, while intrinsic rule-breakers were far from full honesty but closer to justified dishonesty. Signaling induced some intrinsically dishonest people to pretend to be rule-followers for joining the rule-following groups.

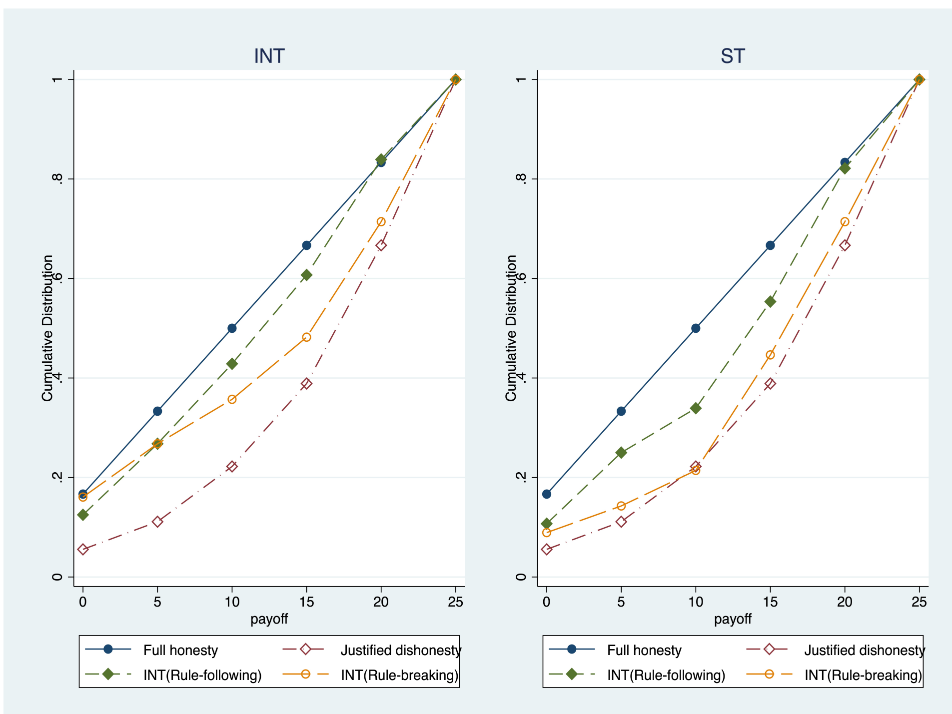


Figure 6: Cumulative distribution of dice game

Table 4: Kolmogorov-Smirnov test for the rule-following and rule-breaking groups

Treatments	H_0	d-value	p-value
INT-RB V.S. FH	INT-RB = FH	0.185	0.012**
INT-RF V.S. FH	INT-RF = FH	0.071	0.592
INT-RB V.S. JD	INT-RB = JD	0.191	0.008***
INT-RF V.S. JD	INT-RF = JD	0.217	0.002***
ST-RB V.S. FH	ST-RB = FH	0.286	0.000***
ST-RF V.S. FH	ST-RF = FH	0.161	0.039**
ST-RB V.S. JD	ST-RB = JD	0.083	0.413
ST-RF V.S. JD	ST-RF = JD	0.173	0.015**

Note: INT and ST are the internalization of norms treatment and the signaling treatment, respectively. RF and RB are the rule-following and rule-breaking groups, respectively. FH is full honesty, and JD is justified dishonesty. * Significance at 10%, ** significance at 5%, *** significance at 1%.

5.3 Repeated Public Goods Game

5.3.1 Treatment Effect in Public Goods Game

Here we discussed how the signaling mechanism impacted cooperation in the repeated public goods game (PGG). Figure 7 shows that participants in the signaling treatment (ST) had an average contribution of 34.4%, which is significantly greater than that of the baseline treatment (BT=30.9%; Wilcoxon-Mann-Whitney test: z=3.315, p=0.001) but insignificantly different from the Internalization of Norms Treatment (INT=35.0%; Wilcoxon-Mann-Whitney test: z=0.065, p=0.949). The proportion of contributions in the Internalization of Norms Treatment (INT) is significantly greater than BT (Wilcoxon-Mann-Whitney test: z=3.155, p=0.002). As shown in Table 5, we also performed an OLS regression on contributions in different treatments. We found, compared with BT, that participants in both INT and ST contributed more to the public account (b=4.106, p=0.006; b=3.507, p=0.016), while no significant difference is found between INT and ST (b=-0.598, p=0.667).

Figure 8 displays contributions over periods by each group and associated standard errors. Contributions declined over periods in all treatments. As Table 6 shows, a regression analysis of mean group contributions in each period provides statistical support for the declines over periods. We employed mixed-effects regression with random effects for each group, and we clustered standard errors at the group level (Kimbrough and Vostroknutov, 2016). As Table 6 shows, the average group contribution has a significant period effect in BT (b=-2.020, p=0.000), indicating that the average contribution in BT decreased significantly over periods. Both INT (b=-2.779, p=0.000) and ST have a significant downtrend (b=-1.769, p=0.000).

Result 3: Compared with the results of the baseline treatment (BT), both the internalization of norms treatment (INT) and the signaling treatment (ST)

showed a higher level of cooperation, and there was no significant difference between these two treatments (INT and ST).

Table 5: OLS regression on contributions in public goods game

Specifications	Dependent variable: Contributions		
	OLS (1)	OLS (2)	OLS (3)
Baseline		-4.106***	-3.507**
		(1.479)	(1.451)
Internalization	4.106***		0.598
	(1.479)		(1.391)
Signaling	3.507**	-0.598	
	(1.451)	(1.391)	
Constant	30.917***	35.022***	34.424***
	(1.085)	(1.004)	(0.962)
F-value	4.43	4.43	4.43
R^2	0.0028	0.0028	0.0028
Observations	3200	3200	3200

Note: N=3200; 4 players×80 groups×10 observations. The dependent variable is contributions in the public goods game. Baseline, internalization, and signaling are dummy variables. Baseline is equal to 1 when it represents the baseline treatment. Internalization is equal to 1 when it represents the internalization of norms treatment. Signaling is equal to 1 when it represents the signaling treatment. Standard error in parentheses. * Significance at 10%, ** significance at 5%, *** significance at 1%.

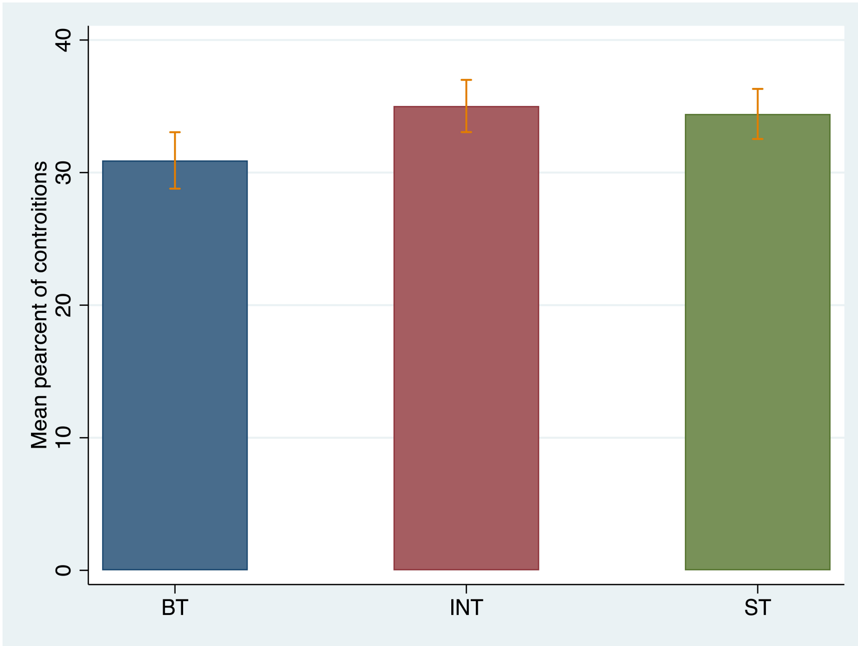


Figure 7: Means of contributions in the public goods game

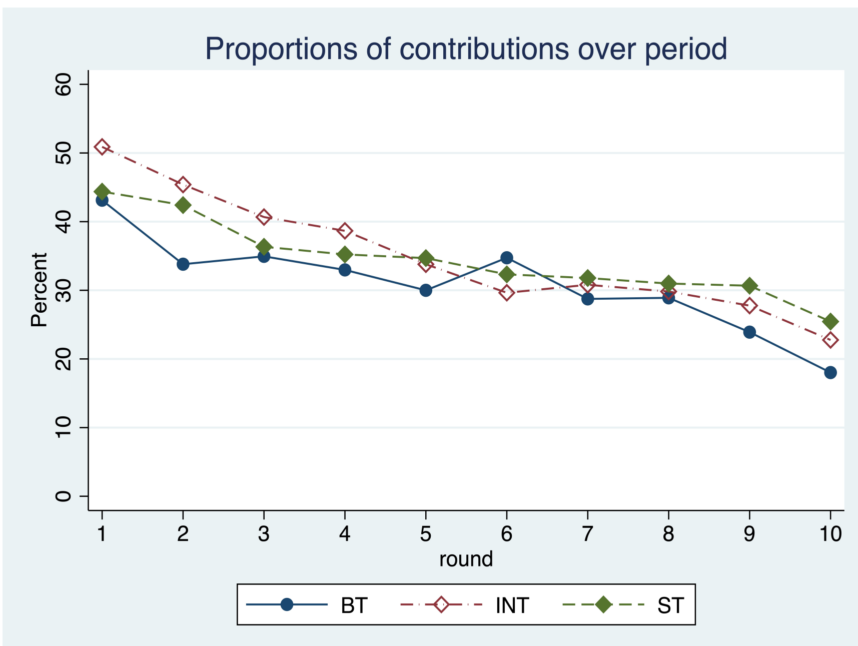


Figure 8: Contribution trends over periods

Table 6: Estimates of group mean PG contributions (out of 20 tokens)

Specifications	Dependent variable: Contributions		
	BT	INT	ST
Period	-2.020***	-2.779***	-1.769**
	(0.394)	(0.371)	(0.286)
Constant	42.028***	50.307***	44.158***
	(3.319)	(3.034)	(2.447)
$Sigma_u$	20.424	21.910	22.682
$Sigma_e$	26.169	24.289	22.392
rho	0.379	0.449	0.506
Groups	24	28	28
Participants	96	112	112
Observations	240	280	280

Note: N=240; 24 groups×10 observations in BT. N=280; 28 groups×10 observations in INT and ST. The dependent variable is the percentage of contributions in the public goods game. * Significance at 10%, ** significance at 5%, *** significance at 1%.

5.3.2 Cooperation in Rule-Following and Rule-Breaking Groups

As Result 2 shows, there seems to be no signaling effect in contributions by comparing ST to INT, which indicates that the signaling failed to improve the level of cooperation. However, it is possible that the signaling mechanism has opposite effects in rule-following and rule-breaking groups, which would lead to an offset of the difference between INT and ST.

Rule-following groups have a higher level of contributions than rule-breaking groups (see Figure 9 and Table 7). Specifically, in INT, participants in rule-following groups made an average contribution of 40.27% to the public account, which is significantly higher than those in rule-breaking groups (29.78%) (Wilcoxon-Mann-Whitney test: z=5.359, p=0.000; OLS: b=10.491, p=0.000). Similarly, in ST, participants in rule-following groups on average contributed 46.13% of their endowments to the public account, which is significantly more than those in rule-breaking groups (22.72%) (Wilcoxon-Mann-Whitney test: z=12.267, p=0.000; OLS: b=23.402, p=0.000).

The signaling mechanism had a positive effect by improving cooperation in the rule-following groups; specifically, ST had a greater level of contributions than INT (Wilcoxon-Mann-Whitney test: z=3.236, p=0.001; OLS: b=5.857, p=0.000). However, we also found a significantly negative effect of the signaling on contributions in rule-breaking groups. Contributions in ST (22.72%) were less than those in INT (29.78%) (Wilcoxon-Mann-Whitney test: z=3.035, p=0.002; OLS: b=-7.054, p=0.000). The heterogeneous effect on contributions in rule-following and rule-breaking groups of the ST produced an offset that is, in total, not different from INT. Although the signaling induced a disguise, which means that in ST, low-quality signalers joined the rule-following groups, this actually worked to increase contributions. In other words, there must be some intrinsic rule-followers who are crowded out from the rule-following groups and crowded into the rule-breaking groups, such that in ST, the contributions from rule-breaking groups were eventually lower than in the full rule-breaking groups in INT.

Result 4: Rule-followers made more contributions to the public goods game than rule-breakers. The signaling had a positive impact on improving cooperation in rule-following groups, but it had a negative effect in rule-breaking groups.

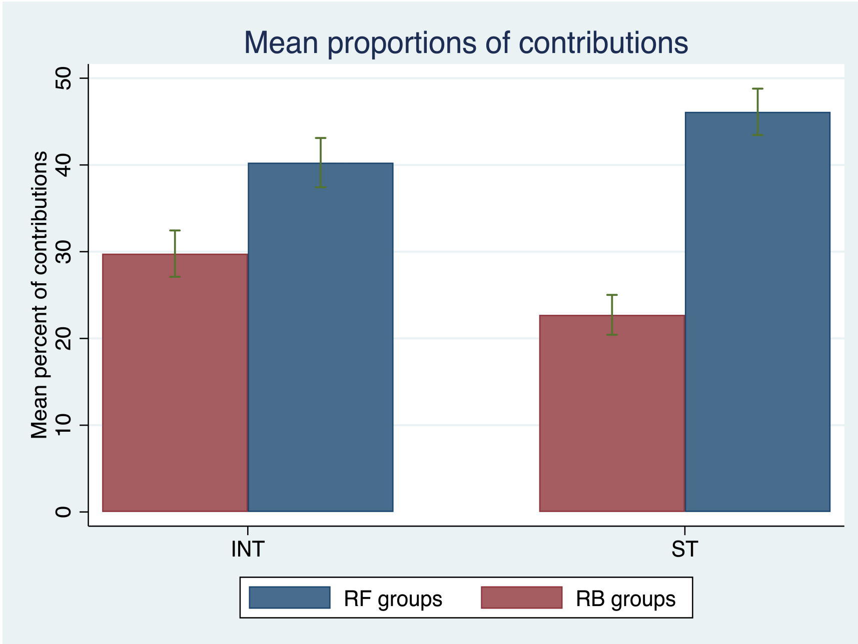


Figure 9: Means of contributions in the rule-following and rule-breaking groups

Table 7: OLS regression for signaling and internalization of norms treatments

Specifications	Dependent variable: Contributions			
	OLS (1)	OLS (2)	OLS (3)	OLS (4)
INT-RF	17.545***	-5.857***	10.491***	
	(1.859)	(1.984)	(1.985)	
INT-RB	7.053***	-16.348***		-10.491***
	(1.794)	(1.923)		(1.985)
ST-RF	23.402***		16.348***	5.857***
	(1.794)		(1.923)	(1.984)
ST-RB		-23.402***	-7.054***	-17.545***
		(1.794)	(1.794)	(1.859)
Constant	22.723***	46.125***	29.777***	40.268***
	(1.169)	(1.359)	(1.360)	(1.445)
F-value	66.82	66.82	66.82	66.82
R^2	0.0760	0.0760	0.0760	0.0760
Observations	2240	2240	2240	2240

Note: In this regression, we compared ST with INT. N=1920; 4 players×48 groups×10 observations. The dependent variable is contributions in the public goods game. INT and ST are the internalization of norms treatment and signaling treatment, respectively. RF is rule-following and RB is rule-breaking groups, respectively. INT-RF is a dummy variable, and it is equal to 1 when it represents the rule-following groups in INT. INT-RB is a dummy variable that is equal to 1 when it represents the rule-breaking groups in INT. ST-RF is a dummy variable, and it is equal to 1 when referring to the rule-following groups in ST. ST-RB is a dummy variable, and it is equal to 1 when referring to the rule-breaking groups in ST. * Significance at 10%, ** significance at 5%, *** significance at 1%.

The results of the INT are consistent with Kimbrough and Vostroknutov (2016) in which the norm in the repeated PGG was one of conditional cooperation. To illustrate the norm of conditional cooperation, we assume that if the RF game captures other-regarding, instead of norm dependence, then the contributions of rule-breakers should systematically differ over time from those of rule-followers, even in the absence of sorting (Kimbrough and Vostroknutov, 2016). As Table 8 shows, there is significant and positive correlation between the RF game and contributions in the PGG. This is similar to the result of Kimbrough and Vostroknutov (2016), but the positive and significant coefficient of mean other $contribution_{t-1}$ and the insignificant coefficient of the interaction indicates that the norm of conditional cooperation is the main motivation of contributions in both the rule-following and rule-breaking groups. This finding is different from the result of Kimbrough and Vostroknutov (2016) in which contributions made by individuals who had a higher level of following social norms in the

RF task were more responsive to the contributions made by others. We conclude that the norm of conditional cooperation is the main motivation of contributions by rule-followers and rule-breakers in the PGG.

Result 5: The individual rule-following behavior correlates with individual contributions. Rule-followers and rule-breakers conditionally contributed to the public goods game on the contributions of other members.

Table 8: Mixed-effect estimates of contributions in PG, baseline treatment

PG $contribution_t$	Coef.	Std. Err.	z-value	$P > \|z\|$
(Intercept)	20.157	4.318	4.67	0.000***
Period	-1.444	0.406	-3.56	0.000***
RF game	0.144	0.073	1.96	0.050*
Mean other $contribution_{t-1}$	3.777	0.072	5.21	0.000***
RF game $\times$ Mean other $contribution_{t-1}$	0.0001	0.001	0.10	0.919

Note: The RF game represents the rule-following game, which is scored based on the number of rubbish balls put in the trash can (Box A). N=800; 20 groups$\times$4 participants$\times$10 periods. Standard error clustered at the group level. * Significance at 10%, ** significance at 5%, *** significance at 1%.

5.3.3 The Heterogeneity of Signaling Effects

Figure 10 depicts the trends of mean contributions in rule-following and rule-breaking groups. We performed a mixed-effect regression on group mean PG contributions in INT and ST with clustering standard errors at the group level (Kimbrough and Vostroknutov, 2016). We regressed group mean contributions on an intercept, a period trend, a rule-following group dummy and a period$\times$rule-following dummy.

In INT, participants had no idea about other members' propensity to follow the norm $\phi_{RF,-i}$, so they did not have any information on the propensity to conditionally cooperate $\phi_{PGG,-i}$ at the beginning of the repeated PGG. In other words, there should be no difference in $contribution_1$ between rule-following and rule-breaking groups. However, participants spend time deducing the norm in by observing the mean contributions of others in $period_t$ ($t \geq 2$), so contributions should separate over time between rule-following and rule-breaking groups. We tested $contribution_1$ between rule-following and rule-breaking groups with the Wilcoxon-rank-sum test. We found no significant difference in $contribution_1$ (Wilcoxon-Mann-Whitney test: z=0.386, p=0.699). Then, we examined whether there exists a separation trend between rule-following and rule-breaking groups with the mixed-effect estimates of group mean PG contributions (see Table 9). As we saw a positive and significant coefficient, this indicates that the slope of contributions over time in rule-following groups is significantly greater than that of rule-breaking groups (b=2.088, p=0.017), and, because $contribution_1$ is not different between two cohort groups, we can conclude that there is a significant separation in contributions over periods in INT.

The $contribution_1$ in rule-following and rule-breaking groups was 52.1% and 49.6%, respectively, while the last-period contributions ($contribution_{10}$) are 31.7% and 13.8%, respectively. The effect was partly offset among rule-followers, which corresponds with the result of Kimbrough and Vostroknutov (2016), but the decay also appears in the rule-following groups, which is contrast to K&V. Yet, in that study, they only informed participants to play "several" rounds of the repeated PGG, so participants did not have an expectation of when the game would end; this allowed cooperation to be sustained at a high level instead of declining over periods. Contrary to INT, the $contribution_1$ under the signaling mechanism was mainly impacted by the belief about others' propensity to follow social norms ($\phi_{RF,-i}$) in the RF game. Due to the difference in ($\phi_{RF,-i}$), the $contribution_1$ should be different between rule-following and rule-breaking groups. The $contribution_1$ in rule-following groups is 53.7%, whereas that of rule-breaking groups was 35.1%, representing a significant difference in $contribution_1$ between rule-following and rule-breaking groups (b=21.077, p=0.000), which is consistent with Hypothesis 3.

Result 6: There was no difference in first-period contributions between rule-following and rule-breaking groups in the internalization of norms treatment (INT); conversely, under the signaling mechanism, participants in rule-following groups had a higher level of first-period contributions than rule-breaking groups.

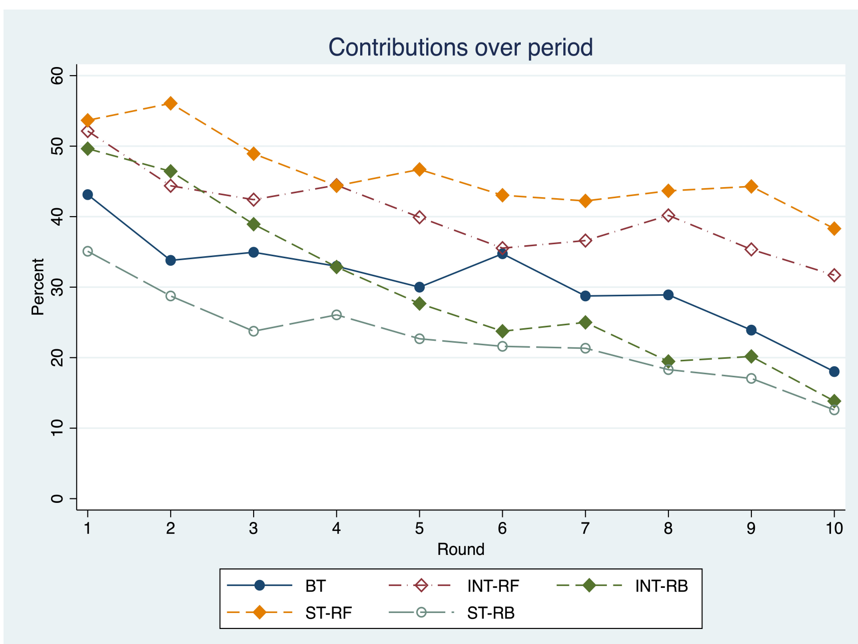


Figure 10: Contribution trends of rule-following and rule-breaking groups

Table 9: Mixed-effect estimates of group mean PG contributions

Specifications	Dependent variable: Contributions	
	INT	ST
Periods	-3.823***	-1.981***
	(0.669)	(0.547)
Rule-following	-0.994	21.077***
	(9.107)	(5.815)
Periods×Rule-following	2.088**	0.423
	(0.872)	(0.710)
Intercept	50.804***	33.619***
	(6.151)	(3.719)
$Sigma_u$	17.473	16.583
$Sigma_e$	12.852	12.348
rho	0.649	0.643
Groups	28	28
Participants	112	112
Observations	280	280

Note: N=280; 28 groups×10 observations in INT and ST. The dependent variable is the percentage of contributions in the public goods game. Rule-following is a dummy variable, and it is equal to 1 when it represents the rule-following behavior in the rule-following game. Periods×Rule-following is an interaction variable. Standard error in parentheses. * Significance at 10%, ** significance at 5%, *** significance at 1%.

In Table 10, for rule-following groups, the first-period contributions in ST is greater than that of INT but it is not significant (b=4.887, p=0.351), and the insignificant coefficient on the interaction between the period and the signaling indicates that there is no significant difference in the slope of declined contributions over periods between INT and ST (b=-0.176, p=0.787). However, in rule-breaking groups, participants in ST contributed less to the PGG than did participants in INT (b=-17.185, p=0.002) at the first-period contribution, and the positive and significant coefficient on the interaction between the period and the signaling indicates that there is a difference in the slope of decline in contributions of rule-breaking groups between ST and INT (b=1.842, p=0.004). In rule-breaking groups, the norm of conditional cooperation made participants defect only if other members defected, which caused contributions to decline quickly, while the belief of the propensity about others $\phi_{PGG,-i}$ to cooperate partly offset the declining trend. Participants in rule-following groups cooperated only if other members cooperated, but there is no significant offset effect in ST. We can conclude that the signaling offset decline in contributions in rule-breaking groups, but the rule-breaking groups of ST had a lower level of contributions over periods than the rule-breaking groups of INT. Conversely, the signaling effect in rule-following groups produced higher level of contributions over periods than in the rule-following groups of INT, but there was no offset effect.

Result 7: The signaling treatment (ST) produced a heterogeneous effect on contributions in the rule-following and rule-breaking groups. The signaling mechanism improved contributions in rule-following groups but decreased contributions in rule-breaking groups. The signaling mechanism, thus, had an offset effect on the decline of contributions over periods in rule-breaking groups rather than rule-following groups.

Table 10: Mixed-effect estimates of contributions by rule-following (rule-breaking) groups

Specifications	Dependent variable: Contribution	
	Rule-following groups	Rule-breaking groups
Periods	-1.735***	-3.823***
	(0.493)	(0.517)
Signaling	4.887	-17.185***
	(5.242)	(5.419)
Periods×Signaling	0.176	1.842***
	(0.654)	(0.639)
Constants	49.809***	50.804***
	(4.137)	(4.445)
$Sigma_u$	22.703	17.963
$Sigma_e$	23.959	22.539
rho	0.473	0.388
Observations	1120	1120

Note: N=1120; 4 players×28 groups×10 periods. Standard errors clustered at the subject level. * Significance at 10%, ** significance at 5%, *** significance at 1%.

6 Conclusion and Discussion

This paper investigates how signaling that one follows social norms impacts cooperation in the repeated public goods game. We found that rule-followers put more rubbish balls in the trash can than rule-breakers, which means that rule-followers indeed have a higher propensity of following social norms than rule-breakers, which is consistent with Kimbrough and Vostroknutov (2016). Under the signaling mechanism, disguising took place in rule-following groups, so there should be a pooling equilibrium in the RF game. Participants in rule-following groups contributed more than those in rule-breaking groups in the first period. However, there was a separation in contributions between rule-following and rule-breaking groups over periods, indicating that participants in rule-following groups sustained a higher level of cooperation than those in rule-breaking groups over time.

The signaling treatment had no less cooperation than the internalization of norms treatment, but the signaling treatment produced heterogeneous effects on contributions in rule-following and rule-breaking groups. In particular, compared with the treatment without the signaling mechanism, signaling improved the cooperation of rule-following groups but decreased that of rule-breaking groups. On the other hand, the signaling treatment had an offset effect on cooperation in the rule-breaking groups but not in the rule-following groups. In the internalization of norms treatment, the declining slope of contributions in the rule-breaking groups was greater than that of the rule-following groups, while in the signaling treatment, there was no significant difference. This means that the signaling mechanism can eliminate the decline in cooperation seen in rule-breaking groups. Although disguising took place in rule-following groups of the signaling treatment, it did not increase but rather decreased the downtrend of contributions over periods. This indicates that disguisers have a belief about others' high propensity to follow norms (the norm in the RF game and conditional cooperation in the repeated PGG), so that they are more willing to cooperate rather than defect. Signaling by following social norms is an effective and cheap way to improve cooperation. Compared with the result of the baseline treatment, the cooperation rate in the signaling treatment was significantly higher, which supports the rationality of Posner's signaling mechanism using social norms.

The implications of this paper have some significance for social cooperation. First, this paper investigates how a signaling mechanism impacts people's cooperation, and it illustrates that following social norms as a signal is a good method to establish cooperation besides altruistic punishment, communication, and other traditional mechanisms. Secondly, this paper provides an effective and low-cost way to improve social cooperation, especially since it is easier for people to obtain information about cooperative attitudes by observing others' actions of following social norms. Finally, we disentangle effects of the strategy motivation and the internalization of social norms by comparing the treatment under the signaling mechanism to treatments without signaling. The heterogeneity of effects from the signaling are explored in rule-following and rule-breaking groups. Although this work uses a laboratory experiment to study how to promote social cooperation, it does not include social contexts, such as social identity and social networks; future work in this area could include these factors.

References

[1] Andreoni J. Cooperation in public-goods experiments: kindness or confusion?[J]. The American Economic Review, 1995: 891-904.

[2] Andreoni J. Why free ride?: Strategies and learning in public goods experiments[J]. Journal of Public Economics, 1988, 37(3): 291-304.

[3] Beersma B, Van Kleef G A. How the grapevine keeps you in line: Gossip increases contributions to the group[J]. Social Psychological and Personality Science, 2011, 2(6): 642-649.

[4] Bereczkei T, Birkas B, Kerekes Z. Altruism towards strangers in need: costly signaling in an industrial society[J]. Evolution and Human Behavior, 2010, 31(2): 95-103.

[5] Bird R B, Power E A. Prosocial signaling and cooperation among Martu hunters[J]. Evolution and Human Behavior, 2015, 36(5): 389-397.

[6] Bochet O, Page T, Putterman L. Communication and punishment in voluntary contribution experiments[J]. Journal of Economic Behavior and Organization, 2006, 60(1): 11-26.

[7] Bohm P. Estimating demand for public goods: An experiment[J]. European Economic Review, 1972, 3(2): 111-130.

[8] Brekke K A, Hauge K E, Lind J T, et al. Playing with the good guys. A public good game with endogenous group formation[J]. Journal of Public Economics, 2011, 95(9-10): 1111-1118.

[9] Cappelen A W, Nielsen U H, Sørensen E Ø, et al. Give and take in dictator games[J]. Economics Letters, 2013, 118(2): 280-283.

[10] Carpenter J P. Punishing free-riders: How group size affects mutual monitoring and the provision of public goods[J]. Games and Economic Behavior, 2007, 60(1): 31-51.

[11] Carvalho J P. Veiling[J]. The Quarterly Journal of Economics, 2013, 128(1): 337-370.

[12] Dawes R M, McTavish J, Shaklee H. Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation[J]. Journal of personality and social psychology, 1977, 35(1): 1.

[13] Dawes R M. Social dilemmas[J]. Annual review of psychology, 1980, 31(1): 169-193.

[14] Fehr E, Fischbacher U, Gächter S. Strong reciprocity, human cooperation, and the enforcement of social norms[J]. Human nature, 2002, 13(1): 1-25.

[15] Fehr E, Gächter S. Altruistic punishment in humans[J]. Nature, 2002, 415(6868): 137.

[16] Fehr E, Gachter S. Cooperation and punishment in public goods experiments[J]. American Economic Review, 2000, 90(4): 980-994.

[17] Fehr E, Gächter S. Do incentive contracts crowd out voluntary cooperation? [J]. 2001.

[18] Feinberg M, Willer R, Schultz M. Gossip and ostracism promote cooperation in groups[J]. Psychological science, 2014, 25(3): 656-664.

[19] Fischbacher U, Föllmi-Heusi F. Lies in disguise—an experimental study on cheating[J]. Journal of the European Economic Association, 2013, 11(3): 525-547.

[20] Fischbacher U, Gächter S, Fehr E. Are people conditionally cooperative? Evidence from a public goods experiment[J]. Economics Letters, 2001, 71(3): 397-404.

[21] Fischbacher U. z-Tree: Zurich toolbox for ready-made economic experiments[J]. Experimental economics, 2007, 10(2): 171-178.

[22] Gächter S, Schulz J F. Intrinsic honesty and the prevalence of rule violations across societies[J]. Nature, 2016, 531(7595): 496.

[23] Gambetta D. Codes of the Underworld[M]//Codes of the Underworld. Princeton University Press, 2011.

[24] Gintis H, Smith E A, Bowles S. Costly signaling and cooperation[J]. Journal of theoretical biology, 2001, 213(1): 103-119.

[25] Grafen A. Biological signals as handicaps[J]. Journal of theoretical biology, 1990, 144(4): 517-546.

[26] Haley K J, Fessler D M T. Nobody's watching?: Subtle cues affect generosity in an anonymous economic game[J]. Evolution and Human behavior, 2005, 26(3): 245-256.

[27] Hauert C, De Monte S, Hofbauer J, et al. Volunteering as red queen mechanism for cooperation in public goods games[J]. Science, 2002, 296(5570): 1129-1132.

[28] Hawkes K, Bliege Bird R. Showing off, handicap signaling, and the evolution of men's work[J]. Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews, 2002, 11(2): 58-67.

[29] Isaac R M, Walker J M. Communication and free-riding behavior: The voluntary contribution mechanism[J]. Economic inquiry, 1988a, 26(4): 585-608.

[30] Isaac R M, Walker J M. Group size effects in public goods provision: The voluntary contributions mechanism[J]. The Quarterly Journal of Economics, 1988b, 103(1): 179-199.

[31] Keser C, Van Winden F. Conditional cooperation and voluntary contributions to public goods[J]. scandinavian Journal of Economics, 2000, 102(1): 23-39.

[32] Kessler J B, Leider S. Norms and contracting[J]. Management Science, 2012, 58(1): 62-77.

[33] Kimbrough E O, Miller J B, Vostroknutov A. Norms, frames and prosocial behavior in games[R]. Working paper, Simon Fraser University, Bocconi University, Maastricht University, 2014.

[34] Kimbrough E O, Vostroknutov A. A portable method of eliciting respect for social norms[J]. Economics Letters, 2018, 168: 147-150.

[35] Kimbrough E O, Vostroknutov A. Norms make preferences social[J]. Journal of the European Economic Association, 2016, 14(3): 608-638.

[36] Krupka E L, Weber R A. Identifying social norms using coordination games: Why does dictator game sharing vary?[J]. Journal of the European Economic Association, 2013, 11(3): 495- 524.

[37] Krupka E, Weber R A. The focusing and informational effects of norms on pro-social behavior[J]. Journal of Economic Psychology, 2009, 30(3): 307-320.

[38] Ledyard J O. Public goods: A survey of experimental research[J]. 1994.

[39] López-Pérez R. Aversion to norm-breaking: A model[J]. Games and Economic behavior, 2008, 64(1): 237-267.

[40] Masclet D, Noussair C, Tucker S, et al. Monetary and nonmonetary punishment in the voluntary contributions mechanism[J]. American Economic Review, 2003, 93(1): 366-380.

[41] Normann H T, Wallace B. The Impact of the termination rule in cooperation experiments[J]. Royal Holloway, University of London: Discussion papers in economics, 2004.

[42] Ockenfels A, Weimann J. Types and patterns: an experimental East-West-German comparison of cooperation and solidarity[J]. Journal of public economics, 1999, 71(2): 275-287.

[43] Posner E A. Law and social norms[M]. Harvard university press, 2009.

[44] Rege M, Telle K. The impact of social approval and framing on cooperation in public good situations[J]. Journal of Public Economics, 2004, 88(7-8): 1625-1644.

[45] Roberts G. Competitive altruism: from reciprocity to the handicap principle[J]. Proceedings of the Royal Society of London. Series B: Biological Sciences, 1998, 265(1394): 427-431.

[46] Saad G. The evolutionary bases of consumption[M]. Psychology Press, 2007.

[47] Santos F C, Santos M D, Pacheco J M. Social diversity promotes the emergence of cooperation in public goods games[J]. Nature, 2008, 454(7201): 213.

[48] Sefton M, Shupp R, Walker J M. The effect of rewards and sanctions in the provision of public goods[J]. Economic inquiry, 2007, 45(4): 671-690.

[49] Smith E A, Bird R L B. Turtle hunting and tombstone opening: Public generosity as costly signaling[J]. Evolution and Human Behavior, 2000, 21(4): 245-261.

[50] Soler M. Costly signaling, ritual and cooperation: Evidence from Candomblé, an Afro-Brazilian religion[J]. Evolution and Human Behavior, 2012, 33(4): 346-356.

[51] Spence A M. Market signaling: Informational transfer in hiring and related screening processes[M]. Harvard Univ Pr, 1974.

[52] Szolnoki A, Perc M. Reward and cooperation in the spatial public goods game[J]. EPL (Europhysics Letters), 2010, 92(3): 38003.

[53] Wilson M, Daly M. Competitiveness, risk taking, and violence: The young male syndrome[J]. Ethology and sociobiology, 1985, 6(1): 59-73.

[54] Wright J. Altruism as a signal: Zahavi's alternative to kin selection and reciprocity[J]. Journal of Avian Biology, 1999: 108-115.

[55] Wu J, Balliet D, Van Lange P A M. Reputation, gossip, and human cooperation[J]. Social and Personality Psychology Compass, 2016, 10(6): 350-364.

[56] Yamagishi T. The provision of a sanctioning system as a public good[J]. Journal of Personality and Social Psychology, 1986, 51(1): 110.

[57] Zahavi A. The cost of honesty (further remarks on the handicap principle)[J]. Journal of Theoretical Biology, 1977, 67(3): 603-605.