# Information Disclosure and Cooperation in a Finitely-repeated Dilemma: Experimental Evidence

Kamei, Kenju

Durham University

25 September 2016

# Information Disclosure and Cooperation in a Finitely repeated Dilemma: Experimental Evidence

Kenju Kamei*

Department of Economics and Finance, University of Durham,
Mill Hill Lane, DH1 3LB, United Kingdom.

This version: September 2017

**Abstract:** A large volume of past research has suggested that making information on people's past behaviors visible to others may enhance cooperation in finitely repeated environments. But, do people cooperate with randomly-matched peers by voluntarily revealing their past when they have an option to conceal it? This paper experimentally shows that while voluntary information disclosure does help strengthen cooperation in a random-matching environment, such disclosure does not have effects if it involves a cost because a large fraction of people does not reveal their past. The data also shows that, when subjects can choose an environment with a reputation mechanism or one without it, a stable number of subjects join the reputation community (where their past is revealed) and cooperate with others, especially when the sorting involves a cost. However, some subjects stay away from the reputation community and fail to cooperate with peers in the 'anonymous' community (where their past is not revealed). When there is no cost for sorting process, the reputation mechanism's high efficiency may decrease because some subjects frequently switch between the two communities and attempt to exploit cooperators, although they could cooperate with others if they were forced to disclose without opportunities to choose.

*Keywords*: experiment, cooperation, finitely repeated dilemma, repeated games, reputation

*JEL code*: C92, D74, D83

**Acknowledgement**: This project was supported by research grants from the Telecommunications Advancement Foundation and from the Murata Science Foundation. Durham University Business School provides additional funding. The author thanks the IT team at the Durham University Business School and Paudie Lynch (the IT manager) for help in managing the ORSEE recruiting system and computers when he conducted the experiment.

* Email: kenju.kamei@gmail.com, kenju.kamei@durham.ac.uk.

## 1. Introduction

How people can successfully cooperate in dilemma situations is one of the oldest and the most important questions in economics. One of the actively studied dilemmas in the last few decades by scholars, including social scientists and biologists, is two-person dilemmas. In a two-person dilemma game, the total payoff amount of two players is maximized when they both choose cooperation. However, under the assumption that players are self-interested and that they believe all of their peers are also self-interested, the Pareto-efficient outcome cannot be achieved if it is only finitely repeated. This is because defecting on one's partner results in a higher payoff for the defector, no matter what actions the partner takes.

Scholars have long been interested in how to sustain cooperation in such finitely repeated two-person dilemmas with random matching. As discussed in the seminal work by Kreps *et al.* (1982), mutual cooperation is theoretically possible if people believe that their peers are not selfish or they believe that some of their peers believe that not everyone is selfish. Nevertheless, it is empirically known that although some people attempt to cooperate in earlier periods of finitely repeated dilemma games, people's level of cooperation steadily declines over time if no additional institutions are available (e.g., Dal Bó and Dal Bó, 2014; Kamei, 2016a).[1] However, a large volume of studies have proposed that reputation, or information on people's past behaviors, may help people resolve dilemmas and achieve mutual cooperation in various setups (e.g., Engelmann and Fischbacher, 2009; Bolton *et al.*, 2004, 2005; Kamei and Putterman, 2017; Nowak *et al.*, 2000; Milinski *et al.*, 2002).[2] Research so far has indicated that three particular factors come into play in a reputation mechanism. First, the presence of a reputation mechanism may give people material incentives to build a cooperative reputation and to cooperate (e.g., Andreoni and Miller, 1993; Kamei and Putterman, 2017; Bolton *et al.*, 2004, 2005; Engelmann and Fischbacher, 2009). Maintaining a mutually cooperative relationship is more materially beneficial than exploiting counterparts and then being trapped in a mutual defection if they meet

[1] Also see Ledyard (1995) and Chaudhuri (2011).
[2] The impact of exogenously given information on boosting cooperation has also been demonstrated in infinitely repeated dilemma game experiments with random matching (e.g., Camera and Casari, 2009; Kamei, 2017).

more than once. Thus, a selfish individual may mimic the behavior of cooperative individuals and strategically build a reputation as a highly cooperative person in order to establish mutually cooperative relationships with others. Second, some individuals may receive psychological satisfaction from mutual cooperation, explained by concepts such as altruism, inequity aversion and direct reciprocity, while they are averse to being exploited by others (see Fehr and Schmidt (2006) and Sobel (2005) for a survey). As information on the past may serve as a signal that they will cooperate or defect with their peers in the future, reputation mechanisms can act as coordination devices for people to fulfill their desires to cooperate. Lastly, some individuals may exhibit indirect reciprocity even towards strangers, as shown by Engelmann and Fischbacher (2009), Seinen and Schram (2006), Nowak and Sigmund (1998a, b), Wedekind and Milinski (2000), and Wedekind and Braithwaite (2002). Because pro-social behavior can be rewarded by strangers, individuals – even selfish ones, may have a sufficient incentive to cooperate when their behavior can be seen by others in the future. Nevertheless, do people in fact choose to disclose their past in a random-matching community and voluntarily cooperate with each other if there is an alternative choice to hide it? Moreover, if subjects' disclosure decisions determine with whom they interact with, i.e., if there is sorting through disclosure in that disclosers are matched with disclosers and non-disclosers are matched with non-disclosers, how do subjects' disclosure decisions change in two-person dilemmas with random matching? Although there is a rich body of studies on the role of reputation institutions ('exogenously' given to subjects) in the evolution of cooperation, little attention has been paid to the possibility where people create reputation communities through voluntary disclosure of information in finitely repeated dilemma situations, or to people's preferences between an environment with and without a reputation mechanism in dilemma situations.

There are many real-world situations in which individuals make information disclosure decisions in random-matching environments. For instance, in some countries, such as China, people's option to voluntarily disclose their reputational information built on online platforms, so-called 'social credit', has emerged and users are encouraged to reveal such scores to potential interaction persons (e.g., Hatton 2015).[3] Another example is companies' disclosure of accounting and/or corporate information in annual reports: some companies voluntarily disclose more detailed information than required by the government. When a firm is looking for a business partner, they may encounter some firms that voluntarily disclose more precise corporate or reputational information than others. Situations where individuals make sorting decisions have also substantially increased in our modern societies. Examples include online-based interactions, such as emerging businesses based on the sharing economy and dating services. For instance, users can choose whether to use (i) Uber services, where a two-way rating and feedback system is present, or (ii) private cruising taxis available on street, for which no information about the reputation of their services is available, when getting a ride to an airport. Using the Uber services may be costly because users need to install the application and also register or update their profiles. Although there is no such cost involved, interactions like option (ii) create more anonymous interactions across users. What fractions of people prefer option (i), instead of option (ii)? Does users' selection of an environment affect their opportunistic behaviors (e.g., a driver treats her passenger poorly or a passenger behaves arrogantly)?[4] Similarly, when deciding on a residential location, one may consider choosing to relocate to an area with well-functioning public monitoring. But in order for a person to become a member of a community with such reputation mechanisms, the new person must become known by other members of the community. For this purpose the person will need to incur some cost (e.g., time) to connect with others and introduce himself (background, personality, etc.) to the community.[5]

[3] On social networks, while some users use their real name and disclose their background, other users only use anonymous usernames. Their disclosure of background information may also be related to their willingness to show their reputation and to interact with like-minded others.

[4] For another example, singles may seek to join an online dating service where the users can see other users' detailed background information (e.g., incomes). These services, however, often require users to pay membership fees or require them to spend some time registering, ask them to submit supporting documents to verify the information in some platforms, and/or updating their profiles that will be made available to other users. Thus, joining in a network with people's personal information is often not cost-free. By contrast, there are usually alternative networks that do not require users to pay or spend time in submitting their detailed information.

[5] Unlike transactions in online platforms or a person's relocation to a new place just discussed, in small-scale economies, such as a close-knit residential community without sorting, news related to residents spread across the community without anyone making an effort. Such situations can be modeled by assuming that information automatically spreads among those involved without a cost; and a person needs to spend a cost if s/he oppositely *does not* want to disclose some information. This situation was recently studied by Kamei (2017).

According to the experimental literature on repeated dilemmas, cooperation may evolve through voluntary information disclosure or through sorting in a finitely repeated environment with random matching. First, past studies on repeated supergames suggest that people indeed prefer to cooperate with each other; and they can achieve a high level of cooperation in earlier rounds of interactions if they repeat supergames under some conditions (e.g., Selten and Stoecker [1986], Andreoni and Miller [1993], Embrey *et al.* [forthcoming], Kamei and Putterman [2017]).[6] Our study does not let subjects repeat finitely repeated dilemma games since we are interested in the impact of information disclosure or sorting without learning effects across supergames (see Section 2). Moreover, the quantity of information depends on subjects' decisions to disclose. Nevertheless, information disclosure or sorting process may serve as coordination devices, and thus people may be able to achieve a high level of cooperation as an outcome of rational behavior (e.g., Tyran and Feld, 2006; Duffy and Feltovich 2002, 2006).

We conducted a finitely repeated two-person public goods game experiment where subjects have an option to disclose their last-period contribution amounts to their matched partners in every period. There are two dimensions in the experimental design. The first dimension is the matching protocol: whether subjects are randomly matched with each other in every period irrespective of their disclosure decision, or subjects are in principle assured that a discloser will be matched with another randomly-selected discloser and likewise a non-discloser will be matched with another non-discloser in each period. The latter means that a subject can sort into a community with a reputation mechanism. The second dimension is the disclosure cost: whether disclosure/sorting is costly or cost-free. Thus, our experimental design is a $2 \times 2$ design.

With this setup, theoretical analyses suggest a possibility of rational cooperation. For instance, we illustrate in Section 3 that it would be materially beneficial for even a selfish player to strategically mimic the behavior of conditional cooperators if there are sufficiently large number of conditional cooperative disclosers in a given community.

[6] The past studies nevertheless found that earlier unraveling of cooperation is also observed with experiences in later supergames

Our experiment showed that information endogenously disclosed indeed helps increase the level of cooperation, compared with when no information is provided. However, the effectiveness depends on the presence of disclosure cost, as subjects are reluctant to disclose their past if it involves a cost and accordingly the quantity of information stays at a low level. Nevertheless, whether disclosure was made costly or for free, the average contributions declined gradually and steadily from period to period.

However, our data showed that if subjects are given an opportunity to join the reputation mechanism *by paying a cost*, they are able to achieve a higher level of cooperation in a community with the reputation mechanism, compared with the random-matching environment. The same does not hold for the treatment where the sorting is free. Nevertheless, the efficiency of the costly-sorting treatment is not significantly different from that of voluntary information disclosure without disclosure cost and sorting, because an inefficient anonymous community also emerges in the former treatment. Detailed analyses revealed that subjects are able to cooperate with each other at high levels if all subjects *are forced to* reveal their past; however, given an option to choose between disclosing and not disclosing, some subjects decide not to disclose and then to behave opportunistically. This paper discusses driving forces behind these results.

The main contributions of our paper to the literature are two-fold. First, this paper contributes to a growing body of experimental literature on the evolution of cooperation in finitely repeated dilemma games, by providing evidence that rational cooperation may emerge through people's voluntary disclosure of their past behavior alone. Second, we show that the level of cooperation is still modest with voluntary information disclosure only due to the prevalence of the non-negligible fraction of non-disclosers, but when people have a costly self-selection opportunity between communities with and without the reputation mechanism, cooperation can be sustained at a high level in the community with reputation, although a sizable anonymous community with low efficiency also emerges.

The rest of the paper proceeds as follows: Section 2 summarizes the experimental design, Section 3 discusses possible subject's behaviors and hypotheses of our study, Section 4 reports results and Section 5 concludes.

## 2. Experimental Design

Our study is based on an incentivized laboratory experiment. The design frame used is a two-person public good game (e.g., Ledyard, 1995; Chaudhuri, 2011). The experiments consist of 20 periods of interactions. In each period, each subject is paired with another subject, is given an endowment of 20 ECUs (experimental currency units) and simultaneously decides how many ECUs they wish to contribute for their pair's joint account.

Each period, except period 1, consists of two stages (Figure 1). In the first stage of a given period $t \in \{2, 3, \ldots, 20\}$, subjects decides whether to disclose their period $t - 1$ allocation decisions to their current partners matched for period *t*. We vary the treatments by two dimensions (Table 1). The first dimension is the size of disclosure cost: either the disclosure is free or one ECU is charged. In costly-disclosing treatments, one ECU is deducted at the end of a given period (a subject has 20 ECUs in her allocation-decision stage even when she decides to costly disclose her last-period allocation amount). The second dimension is the matching protocol: one that a discloser is assured that he or she is matched with another discloser, or the other that each subject is randomly matched with another subject, regardless of disclosure decision (Section 2.1). Our design is therefore a 2 × 2 factorial design. The four treatments are named as the "Costly Sorting" treatment, abbreviated as the C-Soring treatment, "Free Sorting" treatment, abbreviated as the F-Soring treatment, the "Costly Disclosure, Random Matching" treatment, abbreviated as the C-RM treatment, and the "Free Disclosure, Random Matching" treatment, abbreviated as the F-RM treatment. (We also conducted three additional treatments, as will be discussed in Section 4, for better controls to study the impact of sorting and disclosure.)

### *2.1. Matching Protocol*

In period 1, each subject is randomly matched with another subject without making any disclosure or sorting decisions; and then plays the public goods game with the matched subject.

As mentioned, each period after period 1 consists of two stages (Figure 1). In the two sorting treatments (F-Sorting, C-Sorting), subjects in period $t \in \{2, 3, \ldots, 20\}$ decides whether to disclose their period $t - 1$ allocation amounts to join the 'reputation' community in period $t$. If a subject chooses to disclose her previous behavior, she is randomly matched with another subject that likewise chose to disclose his last-period contribution amount. The disclosing cost is one ECU in the C-Sorting and is free in the F-Sorting treatment. Alternatively, a subject could join a community of non-disclosers in period $t$, which we call the 'anonymous' community, by choosing not to disclose her allocation amount in the previous period. In that case, her previous contribution amount would not be informed to her current matched partner (and one ECU is not deducted at the end of period $t$ in the C-Sorting treatment). She is then randomly matched with another non-discloser; and the partner is only informed that he is randomly matched with another person in the community of non-disclosers.[7]

In the two random-matching treatments (C-RM and F-RM), each subject in period $t \in \{2, 3, \ldots, 20\}$ decides whether to disclose his or her period $t - 1$ allocation amount in the period $t$ interaction as in the C-Sorting and F-Sorting treatments, respectively. However, in these two treatments, each subject is randomly matched with another subject, regardless of their disclosure decision. The consequences of subjects' disclosure decisions, except the matching procedures, are the same as those in the sorting treatments. That is, when a subject decides to disclose in period $t$, the subject's matched partner in that period is informed of her period $t - 1$ allocation decision; and the partner is not given this information when the subject decides not to disclose.

[7] If the number of disclosers (non-disclosers) is an odd number, one discloser is randomly matched with a non-discloser. This event happened only around 9.3% of the paring in the C-Sorting treatment and 6.9% of it in the F-Sorting treatment. Further data analyses (Section 4) indicate that the paper's findings are robust, regardless of whether we use data of pairs consisting of pairs with the same preferences (two-disclosers pairs and two-non-disclosers pairs) only or all data.

There are no subject identification numbers provided during the entire experiment. Thus, the only way that subjects could form reputation is through their disclosure (sorting) decisions in the C-RM and F-RM (C-Sorting and F-Sorting) treatments.

*2.2. Allocation Decisions*

Each matched pair in every period plays a two-person linear public goods game. Only integers between 0 and 20 are allowed for their contribution amounts. The payoff consequences are as follows: a subject receives one ECU for each ECU she allocates to her private account. By contrast, if she contributes one ECU to the joint account, she and her partner each receive 0.8 ECUs, which is less than one ECU, from the joint account. However, the total group payoff is maximized when two subjects in a pair both contribute all of their endowments to the joint account ($0.8 \times 2 = 1.6 > 1.0$). Suppose that subject $i$ contributes $c_{i,t}$ to her joint account in period $t$. Then, subject $i$ obtains the following payoff in period $t$:

$$\pi_{i,t} = 20 - c_{i,t} + .8 \cdot (c_{i,t} + c_{j,t}). \quad (1)$$

Here, subject $j$ is subject $i$'s matched person in period $t$ and $c_{j,t}$ is subject $j$'s contribution to the joint account in period $t$.

*2.3. Conditional Contribution Schedule and Beliefs*

We include two kinds of additional tasks in order to explore driving forces behind subjects' reputation building behaviors in the C-RM and F-RM treatments and subjects' community formation in the C-Sorting and F-Sorting treatments as supplementary analyses. The first additional task is elicitation of beliefs. First, in each allocation-decision stage (period 1 and second stage of period $t \in \{2, 3, \ldots, 20\}$) of all treatments, subjects are asked about their beliefs on their matched partner's contribution amount in a given period. We note that in the elicitation stage, subjects are aware of their partners' current-period disclosure decisions and the partners' last-period contribution behaviors in case they select to disclose. Second, subjects in the C-RM and F-RM treatments are also asked to answer the expected number of disclosers (except themselves) in period $t$. These elicitation tasks were not incentivized because our first priority is on subjects' actual behaviors and also because incentivized elicitation may affect subjects' actual contribution behaviors (see Gächter and Renner 2010, for example).[8,9]

Second, we elicit subjects' cooperation types using Fischbacher *et al.* (2001). Specifically, each subject is asked to answer how many ECUs they wish to allocate to their group, conditional on each of the other group members' average contributions. This task is incentivized. The details of the procedure for this part are provided in the online Appendix A. This task is conducted before the 20 periods of the finitely repeated public goods game. However, subjects are informed of the outcomes of this task only after they complete the 20 periods of the public goods games to minimize the effects of this task on their behaviors. In addition, group composition is randomly changed between this elicitation task and the 20 periods of repeated dilemmas. The elicited conditional contribution preferences are used to examine whether cooperation-oriented types are more likely to disclose their past behavior and how selfish types strategically build reputations.

### *2.4. Experimental Procedure*

All experiments except the instructions were programmed using z-tree (Fischbacher, 2007). Eligible subjects were sent solicitation messages via ORSEE (Online Recruitment System for Economic Experiments) developed by Greiner (2015); and subjects voluntarily registered for and participated in the experiment. At the onset of the experiment, neutrally framed instructions for the conditional contribution task were handed out to subjects and read aloud by the experimenter. Once the task for eliciting cooperation types was over, instructions for the main part of the experiment (20 periods of public goods games), which are likewise neutrally framed, were distributed and read aloud. Control questions were included in each set of the instructions so as to check the subjects' understanding of the experiment. Communication was prohibited

[8] In the experiment, this task was only included in the instructions shown on subjects' computer screens, not in the hard copy of instructions distributed to subjects, in order to avoid making this task salient.

[9] Gächter and Renner (2010) experimentally show that subjects' contribution behavior is not affected if belief elicitation is not incentivized, but it is affected if belief elicitation is incentivized. They also show that the difference in belief accuracy is relatively small between when it is incentivized and when it is not. A possible way to incentivize beliefs is, for example, to randomly select some periods for payments based on belief accuracy and the other for payments based on the actual contribution behaviors. We did not employ this method because our first priority is on actual contribution behavior and also because it could make the experiment too complex.

during the entire experiment. Subjects were also asked to switch off any electronic devices (e.g., mobile phones) during the experiment. Subjects were privately paid based on their accumulated ECUs (40 ECUs are exchanged for £1 of real money) at the end of the experiment.

A total of 16 sessions – four sessions for each treatment – were conducted. As explained in Section 4, another ten sessions were also conducted as additional treatments with the same subject pool and recruiting procedure mentioned above to identify the effects of information disclosure and sorting more in details. A total of 292 Durham University students participated in the experiment from August 2015 through August in 2017.[10] No subjects participated in more than one session. The average payment (including show-up fee of £3) was £15.61 with a standard deviation of £1.06. The average duration of the experiment (including payment to the subjects) was around 90 minutes.

## 3. Theoretical Considerations and Discussions

Standard theory, based on agents' selfishness and the common knowledge of rationality, provides a point prediction in our environment because the MPCR is 0.8 (see Eq. (1)). That is, contributing zero ECUs to the joint account is a strictly dominant strategy for each subject in any period ($\partial \pi_{i,t}/\partial c_{i,t} = -.2 < 0$). Therefore, with the logic of backward induction, each subject would contribute nothing to the joint account in every period, assuming that they believe other subjects would always choose defection without considering their opponents' reputations. Considering the peers' uniform full free-riding behavior, no one would costly disclose their past towards the matched person in any period in the C-RM treatment; and likewise no one would costly sort into the reputation community in the C-Sorting treatment, because they would want to save a disclosure/sorting cost. Disclosure or sorting decisions do not affect subjects' payoffs in the F-RM and F-Sorting treatments, respectively, because these actions can be taken for free and their peers would select defection always; hence, subjects would randomly decide whether to disclose and sort in the two treatments.

[10] The session size was 12 subjects in all sessions, except one session each in the C-Sorting, F-RM, F-Sorting, Baseline and C-Sorting-N treatments. The session size was eight for these five sessions.

HYPOTHESIS 1: *Standard Theory Prediction*.

*(a) No one costly discloses their last-period contribution behavior in the C-RM treatment. No one costly sorts into the reputation community in the C-Sorting treatment. (b) Disclosure and sorting decisions are randomly made by subjects in the F-RM and F-Sorting treatments, respectively. (c) Subjects contribute nothing to the joint accounts in each period in all treatments.*

A large body of experimental research has partially confirmed Hypothesis 1(c) in finitely repeated dilemma games when there are no institutions such as disclosing information and sorting involved. It demonstrates that although subjects contribute around 40% to 60% of the endowment in the public goods games and around 30% to 40% of subjects choose to cooperate in the prisoner's dilemma games in earlier periods, they decrease the levels of cooperation steadily from period to period (see, e.g., Ledyard [1995] and Chaudhuri [2011] for surveys on public goods games; Andreoni and Miller [1993], Dal Bó and Dal Bó [2014], Dal Bó *et al.* [2010], and Kamei [2016a] for evidence on prisoner's dilemma games).[11]

But on the other hand, theoretically, 'rational' cooperation can be possible even in a finitely repeated dilemma game if we assume that some people believe that some fraction of their peers act on discriminating strategies, such as the tit-for-tat (e.g., Kreps *et al.*, 1982).[12] In fact, in some setups where subjects repeat a finitely repeated dilemma game (supergame), including the ones with partner matching, past research has found that subjects may be able to achieve high levels of cooperation in earlier periods of later supergames (e.g., Selten and Stoecker, 1986; Andreoni and Miller, 1993; Embrey *et al.*, forthcoming; Kamei and Putterman, 2017).[13]

[11] People's difficulty in achieving mutual cooperation with no institutions, due to strategic uncertainty, has also documented for some coordination games with Pareto-ranked equilibria. For instance, subjects' interactions quickly converge to the worst equilibrium outcome in the minimum games (see Devetag and Ortmann [2007] for a survey).
[12] We note that strategies like the 'tit-for-tat' can be evolutionarily stable and can lead to mutual cooperation also in infinitely repeated interactions (e.g., Axelrod and Hamilton, 1981; Axelrod, 1984; Wedekind and Milinski, 1996).
[13] Our study used a random matching protocol. We note that when a random matching protocol is used, without any reputation mechanisms, the average cooperation rate is known to stay at very low rates even if supergames are repeated (e.g., Andreoni and Miller, 1993; Kamei and Putterman, 2017).

Andreoni and Miller (1993) and Kamei and Putterman (2017) also provide evidence that some subjects are non-selfish types and their presence may contribute to the above phenomenon.[14]

Our studies explore the impact of information disclosure and sorting with a standard single-supergame design under a random-matching protocol. With the disclosure (sorting) institutions alone in the C-RM and F-RM (C-Sorting and F-Sorting) treatments, it might be difficult for subjects to foster cooperation norms in communities. This is because subjects in our study can disclose their last-period contribution amounts only, without showing the past history which contains their last-period partner's profiles (e.g., disclosers or non-disclosers, or the levels of cooperation in case that the last-period partners were disclosers). The lack of such second-order information could make it difficult for a community to achieve high cooperation norms, because a subject can be reluctant to contribute a large amount when she is matched with a discloser who has a low level of contributions in his record, even if the partner was matched with, for example, a non-discloser in the previous period. Further, subjects have the reputation mechanism only and do not have any additional stages such as an informal punishment stage. Thus, a subject needs to select a low level of cooperation if she is matched with a low contributor and wants to punish him. However, such a punitive action would lower the subject's own reputation score. Thus, who imposes such punishment could be an issue.

Even with our framework, nevertheless, subjects' rational cooperation behavior as discussed in Kreps *et al.* (1982) may happen, for example, if there is a non-negligible fraction of subjects who act on the so-called 'conditional cooperation strategy' (e.g., Fischbacher *et al.*, 2001; Fischbacher and Gächter, 2010), or if subjects act on a contagious punishment strategy (grim trigger strategy) as discussed in Kandori (1992) in an infinitely repeated setup. The conditional cooperation strategy is similar to the tit-for-tat strategy.[15] Specifically, the

[14] See Reuben and Suetens (2012) also. They used an indefinitely repeated sequential prisoner's dilemma game with strategy method and showed that a non-negligible fraction of second movers select to cooperate if the game does not end in a given period and matched first movers selected cooperation. The prevalence of non-selfish types has also been documented in past studies to elicit preferences (e.g., Fischbacher *et al.*, 2001; Kurzban and Houser, 2005).

[15] Some subjects' conditional cooperation behaviors can be rationalized by assuming that they have interdependent preferences, such as inequity aversion (e.g., Fehr and Schmidt, 1999) and reciprocity (e.g., Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Cox *et al.*, 2007).

conditional cooperation strategy is where subject *m* contributes $x \cdot S_{j,t}$, where $x \in (0,1]$, in period *t* if her pair partner *j*'s state (last-period contribution) is $S_{j,t} = c_{j,t-1}$ and the state is observable; and *m* contributes zero ECUs, otherwise. A large volume of experimental research has found the prevalence of such conditional cooperators (those who act on the conditional cooperation strategy). The steady decline of contributions in a repeated dilemma game with no institutions mentioned earlier can be interpreted that conditional cooperators are discouraged from cooperating by seeing selfish types free ride (e.g., Fischbacher and Gächter, 2010). The contagious punishment strategy in the latter example is a grim-trigger strategy in which a member who sees one instance of free riding starts to free ride from any future partners believing others would do the same, as Kandori (1992) describes that "trust is attached to the community as a whole." We discuss a possibility of rational cooperation with these two strategies now.

*The Presence of Conditional Cooperators and Subjects' Strategic Reputation Building:*

Let us study a possibility of rational cooperation by examining the reputation building behavior of material payoff maximizer *i*, assuming that there exist three types in the population: (a) subjects who choose to hide and act on the "always to defect" strategy (a strategy where players unconditionally select defection), (b) subjects who choose to disclose and act on a conditional cooperation strategy, and (c) subjects who choose to disclose but behave opportunistically (contribute zero to the joint account).

We first consider the C-RM and F-RM treatments. Let us assume $x = 1$ for the sake of simplicity (a perfect conditional cooperator).[16] Suppose that the fraction of disclosers (type (b) or (c) mentioned above) is $p \times 100$ [%] and that that of non-disclosers (type (a)) is $(1-p) \times 100$ [%]; and the fractions of types (b) and (c) among the disclosers are $p_{cc} \times 100$ [%] and $(1-p_{cc}) \times 100$ [%], respectively. Further, suppose that subject *i* correctly anticipates her peers' behaviors and her belief on $p$ and $p_{cc}$ is correct. Lastly, we assume that conditional cooperators contribute $c_1$ in period 1, where no reputational information is available, and up to $\tilde{c}$, where $\tilde{c} > c_1$, in the rest of periods.

[16] The basic implication we obtain for a material payoff maximizer's behaviors does not change even if $x < 1$.

We can show that with this illustrative framework, it would be materially beneficial for $i$ to mimic the behavior of a conditional cooperator by contributing $\tilde{c}$ if $p\cdot p_{cc}$ is sufficiently large. To see this, let us set up the Hamiltonian $H_{t,i}$ as below:

$$H_{t,i} = 20 - c_{i,t} + r\cdot(c_{i,t} + p\cdot p_{cc}\cdot S_{i,t}\cdot 1_{\text{disclose},t}) - f\cdot 1_{\text{disclose},t} + \lambda_{i,t+1}\cdot\Delta S_{i,t}. \quad (2)$$

Here, $r = .8$, $c_{i,t}$ is $i$'s contribution to the joint account, $S_{i,t} = c_{i,t-1}$ is $i$'s state (reputation), $1_{\text{disclose},t}$ is a dummy which equals 1 when $i$ discloses; and 0, otherwise, $f = 1$ (a cost of disclosure), and $\lambda_{i,t+1}$ is the shadow price of a unit of the reputation state in period $t + 1$. Note that $\Delta S_{i,t} = S_{i,t+1} - S_{i,t} = c_{i,t} - S_{i,t}$. Applying the Maximum principle to (2), we find that the optimal controls are bang-bang, characterized by the following conditions (e.g., Sethi and Thompson, 2006):[17]

Condition 1: Material payoff maximizer $i$ chooses to disclose in period $t$ if $S_{i,t} > 1/(p\cdot p_{cc}\cdot r)$ in the C-RM treatment; and if $S_{i,t} \neq 0$ in the F-RM treatment. (3)

Condition 2: Material payoff maximizer $i$ chooses to contribute $\tilde{c}$ in period $t$ if $p\cdot p_{cc} > max\{.25, 1.25/\tilde{c}\}$ in the C-RM treatment; and if $p\cdot p_{cc} > .25$ in the F-RM treatment. (4)

We can make the same derivations for the C-Sorting and F-Sorting treatments. In these treatments, disclosers are matched with each other; and likewise non-disclosers are matched with each other. In this illustrative analysis, for simplicity, let us exclude the small possibility that a discloser is matched with a non-discloser. We then find the following optimal solutions:[18]

Condition 1': Material payoff maximizer $i$ chooses to disclose in period $t$ if $S_{i,t} > 1/(p_{cc}\cdot r)$ in the C-Sorting treatment; and if $S_{i,t} \neq 0$ in the F-Sorting treatment. (5)

Condition 2': Material payoff maximizer $i$ chooses to contribute $\tilde{c}$ in period $t$ if $p_{cc} > max\{.25, 1.25/\tilde{c}\}$ in the C-Sorting treatment; and if $p_{cc} > .25$ in the F-Sorting treatment. (6)

---

[17] The adjoint equation is: $\Delta\lambda_{i,t} = \lambda_{i,t+1} - \lambda_{i,t} = -\partial H_{t,i}/\partial S_{i,t} = -p\cdot p_{cc}\cdot r\cdot 1_{\text{disclose},t} - \lambda_{i,t+1}\cdot(-1)$. In other words, $\lambda_{i,t} = p\cdot p_{cc}\cdot r\cdot 1_{\text{disclose},t}$ (#1). The optimal control for $c$ is bang-bang, because $\partial H_{t,i}/\partial c_{i,t} = -1 + r + \lambda_{t+1}$ does not depend on $c_{i,t}$., which means that $c_{i,t} = \tilde{c}$ if $-1 + r + \lambda_{t+1} > 0$ (#2); and $c_{i,t} = 0$ if $-1 + r + \lambda_{t+1} < 0$. The optimal disclosure decision of subject $i$ is dependent on $S_{i,t}$. From Eq. (2), we find: $i$ discloses if $S_{i,t} > f/(p\cdot p_{cc}\cdot r)$ (#3) in the C-RM treatment; and if $S_{i,t} \neq 0$ (#4) in the F-RM treatment. Conditions (#1), (#2) and (#3) in this footnote suggest that $i$ chooses to contribute $\tilde{c}$ if $p\cdot p_{cc} > max\{.25, 1.25/\tilde{c}\}$ so long as $\tilde{c} > 1$ in the C-RM treatment. Conditions (#1), (#2) and (#4) in this footnote suggest that $i$ chooses to contribute $\tilde{c}$ if $p\cdot p_{cc} > .25$ so long as $\tilde{c} > 0$ in the F-RM treatment.

[18] The Hamiltonian in this case is: $H_{t,i} = 20 - c_{i,t} + r\cdot(c_{i,t} + p_{cc}\cdot S_{i,t}\cdot 1_{\text{disclose,t}}) - f\cdot 1_{\text{disclose,t}} + \lambda_{t+1}\cdot\Delta S_{i,t}$.

We note that when the equality holds in condition (3) or (5), $i$ would be indifferent between disclosing and not disclosing, in a given treatment. Likewise, when the equality holds in condition (4) or (6), $i$ would be indifferent between contributing 0 and $\tilde{c}$ in a given treatment.

These considerations provide hypotheses in our experiment. First, as we expected, Conditions 1 and 1' suggest that $i$ has more incentives to disclose her past in the F-RM and F-Sorting than in the C-RM and C-Sorting treatments, respectively, because disclosure can be made for free in the former treatments. Second and more importantly, these two conditions suggest that because of the sorting mechanism, the incentive to disclose is larger in the C-Sorting than in the C-RM treatment; and therefore $i$ is more likely to choose to disclose (to enter the reputation community) in the former than in the latter treatment unless $p = 1$.[19]

HYPOTHESIS 2: *Incentives to Disclose in the Presence of Conditional Cooperators*.
*(a) The percentages of those who disclose or sort into the reputation community are larger in the F-RM and F-Sorting than in the C-RM and C-Sorting treatments, respectively. (b) The same percentages are higher in the C-Sorting than in the C-RM treatment. (c) Almost all subjects choose to disclose or join the reputation community in the F-RM and F-Sorting treatments.*[20]

In deriving Conditions 1, 2, 1' and 2' (Hypotheses 2 and 3), we have assumed that the belief formed by $i$ would be correct. The accuracy of subjects' belief formation in the C-RM and F-RM treatments can be partially checked with elicited beliefs in the experiment.

The four conditions also have implications for subjects' contribution behaviors. First, Conditions 2 and 2' suggest that material payoff maximizer $i$ has stronger incentives to mimic the behavior of conditional cooperator in the F-RM and F-Sorting than in the C-RM and C-Sorting treatments, respectively, if $\tilde{c} < 5$, because a larger number of conditional cooperators is needed in the treatments with costly disclosure or sorting to compensate for the 1 ECU fee. $\tilde{c} < 5$ is, however, not a strong condition. Second, comparisons between Conditions 1 and 1', and also between Conditions 2 and 2', suggest that $i$ has stronger incentives to strategically build

[19] Notice that $p \cdot p_{cc} < p_{cc}$.
[20] Notice that $i$ chooses to disclose unless $S_{i,t} = 0$ in these two treatments.

reputations as a conditional cooperator in C-Sorting and F-Sorting than in the C-RM and the F-RM treatments, respectively. Note that Conditions 1' and 2' include $p_{cc}$, not $p{\cdot}p_{cc}$, unlike Conditions 1 and 2, because of the sorting mechanism.[21]

HYPOTHESIS 3: *Incentives to Contribute in the Presence of Conditional Cooperators*.[22]

*(a) The negative impact of positive cost: C(F-RM) ≥ C(C-RM), and C(F-Sorting) ≥ C(C-Sorting).*

*(b) The positive impact of sorting: C(C-Sorting) > C(C-FM), and C(F-Sorting) > C(F-RM).*

All of the aforementioned considerations indicate that when $p_{cc}$ or $p{\cdot}p_{cc}$ is sufficiently large, subjects choose to disclose or join the reputation community and then keep high levels of cooperation, only with selfish motives. Thus, as an extreme situation, subjects do not need to exhibit non-standard preferences as discussed in Kreps *et al.* (1982). That is, if a sufficiently large number of subjects believe that a large percentage of their peers will disclose their past and choose (mimic) such conditional cooperation strategy, the beliefs can become self-fulfilling and cooperation may evolve even if all are selfish types.[23]

*Trust Attached to the Community:*

Let us next study a possibility of evolution of cooperation through a contagious punishment strategy (grim-trigger) discussed above. Specifically, assume that all subjects, except subject $i$, have acted on and will continue to act on the following strategy in the C-RM and F-RM treatments: a subject chooses to disclose and contribute at least $\tilde{c}$ in period $t$ if (i) his partner in any period $s \in \{2, 3, \ldots, t-1\}$ disclosed, had a reputation score $S_{j,s}$ with $S_{j,s} \geq \tilde{c}$, and contributed at least $\tilde{c}$ in that period, and (ii) his partner in period $t$ disclosed her last-period behavior and

[21] For example, if $p = .50$, the threshold values of $p_{cc}$ double in the C-RM and F-RM treatments, compared with the C-Sorting and F-Sorting treatments, respectively.

[22] Here, $C(.)$ indicates the average contribution to the joint account in a given treatment.

[23] We note that mutual cooperation may be more likely to be achieved in the C-Sorting and F-Sorting than in the C-RM and F-RM treatments, respectively, even if there are no material incentives to build cooperative reputations. As briefly mentioned earlier, there are past studies on the prevalence of non-selfish human types. It has been shown that people may be able to efficiently cooperate if they are matched with like-minded others with respect to the degree of cooperativeness (e.g., Gunnthorsdottir *et al.*, 2007; Gächter and Thöni, 2010). In order to disentangle subjects' selfish reputation-building motives from non-selfish motives, we classify subjects' cooperation types by using the conditional contribution task (Fischbacher *et al.*, 2001); and we then study the reputation building behavior for each cooperation type (see Section 4).

have a reputation score higher than or equal to $\tilde{c}$; the subject chooses not to disclose and contributes 0, otherwise. This strategy can be thought of as an extreme form of conditional contribution strategy.

In this situation, as illustrated in Appendix D, it is not materially beneficial for *i* to deviate from the same discriminating strategy before period 20 if $\tilde{c}$ is not sufficiently large, and before period 17 even if $\tilde{c}$ = 19 or 20. Numerical calculations indicate that one instance of defection results in quick breakdown of cooperation norms in a community. For instance, when *i* selects to contribute zero in period *t*, the percentage of 0-contributors (except subject *i*) in the community of 12 subjects becomes 95.5% in period $t + 5$; and it is 99.8% in period $t + 6$. Thus, unless their interactions are at very closer to the end, material payoff maximizer *i* would strategically keep contributing $\tilde{c}$ by mimicking their peers' cooperative behaviors. Thus, this theory suggests that cooperation unravels sharply in later periods.

The same analysis can be applied to the C-Sorting and F-Sorting treatments. The timing of deviation by material payoff maximizer *i* depends on the number of subjects in the reputation community. The timing of defection is earlier than the one in the C-RM and F-RM treatments unless the community size is 12, because the smaller the size of community is, the more quickly defection spreads across the community.

The validity of this possibility can be studied by looking at whether a subject stopped cooperating once one of their interaction partners contribute lower amounts to the joint account.

*Related Literature on Signaling and Endogenous Group Formation:*

Lastly, we discuss that the endogenous feature of disclosure or sorting may affect subjects' behaviors, but the impact on the evolution of cooperation is not obvious in our experiment. On the one hand, the disclosure or sorting mechanism could serve as an additional device for subjects to send a signal of future cooperation (e.g., Tyran and Feld, 2006; Duffy and Feltovich, 2002).[24] For an example of public goods games, Tyran and Feld (2006) show that

[24] Giving an opportunity to subjects to send signals also improve coordination among subjects in coordination games (e.g., Cooper *et al.*, 1992; Kamei, 2016b).

collectively selecting a non-deterrent sanction rule in a one-shot public goods dilemma may encourage subjects to contribute more, compared with when the same rule is exogenously imposed, because voting for sanctions can serve as a signal that subjects would cooperate with others. For an example of prisoner's dilemma games, Duffy and Feltovich (2002, 2006) show that a cheap-talk opportunity (where a subject in a pair can send a costless, nonbinding message to his/her pair partner) makes cooperation more likely.

But on the other hand, cooperation may be less likely to evolve even with the sorting mechanism present, according to the literature in endogenous group formation. The literature shows that when subjects are allowed to change their interaction groups at will (i.e., without their members' agreement for moving-in or out, or without threat of exclusion), selfish individuals may attempt to join groups that have cooperators, aiming to exploit them, and as a result cooperation may be difficult to evolve in $N$-person dilemmas, where group size $N > 2$ (e.g., some treatments in Ahn *et al.*, 2008 and 2009; Ehrhart and Keser, 1999).[25,26] These findings may or may not be extended to our experimental setup, however. Our experimental design is different from this strand of past research in that subjects in our design do not engage in an $N$-person dilemma game with all group members ($N$ persons). Instead, subjects who decide to join the reputation or anonymous community are randomly matched with another subject within their community and they then play the *two*-person public goods game with each other. Thus, although such chasing behaviors of malicious individuals may be prevalent, subjects are able to

[25] Cooperation is known to be more likely to evolve in dilemma games with fixed group size, rather than with variable group size, if individuals are provided with both an ability to choose with whom they interact and sufficient information on other players' past behaviors (e.g., Page *et al.*, 2005; Bayer, 2011; Kamei and Putterman, 2017). We note that the evolution of cooperation is likely to be seen in dilemma games with variable group size if entry to a new group requires the group members' agreement (e.g., some treatments in Ahn *et al.*, 2008 and 2009; Charness and Yang, 2014; Gallo and Yan, 2015), or informal sanctioning institutions are present (e.g., Gürerk *et al.*, 2006; Rockenbach and Milinski, 2006; Fehr and Williams, 2013; Nicklisch *et al.*, 2015). In our view, this positive evidence in these studies is partly caused by subject's ability to enforce norms to less cooperative subjects with an agreement rule or a sanctioning institution, without hurting the reputation scores of punishers or high contributors. We also note that the impact of subjects' partner selection on cooperation has also been reported in setups other than dilemma games (see, for example, Bernard *et al.* [2017] for a gift-exchange game).

[26] Past research on sacrifice, which let subjects self-select among environments with different deviation games, finds that subjects achieve a high level of cooperation when they sacrifice the defector gains (e.g., Grimm and Mengel, 2009; Aimone *et al.*, 2013). However, unlike the past studies, subjects in our study are not provided an opportunity to sacrifice deviation gains, but are provided an opportunity to join an environment with the reputation mechanism. Except the presence of the reputation mechanism, not only subjects' set of action choices but also their payoff structure is the same between the reputation and anonymous communities in our study.

adopt tit-for-tat-like discriminating strategies targeted to specific malicious members with the help of the reputation mechanisms, as discussed along with the conditional cooperation strategy above. As such, chasing behaviors by malicious individuals may be less beneficial in our setup, and cooperation may be sustained at a high level in the reputation community of the sorting treatments. Nevertheless, as already discussed, such punishment or protective actions *lower* a member's own reputation. Accordingly, such actions may not be frequently taken. Even if such punishment is initiated by some subjects, there is a chance in which free riding spreads to other members as an epidemic and may quickly destroy the community's cooperation norms.

## 4. Results

We first overview subjects' contribution behaviors and their disclosure/sorting decisions in Section 4.1. We then study the treatment effects of information disclosure and sorting in Section 4.2. We then explore the driving forces behind the observed patterns of the data in Sections 4.3 to 4.7.

### *4.1. Subjects' Disclosure and Sorting Decisions, and their Average Contributions*

We first overview the trends of subjects' disclosure decisions and average contribution amounts in the C-RM and F-RM treatments. Figure 2 reports the trends. The data shows that Hypothesis 1 does not hold. Even when there were no sorting opportunities, a non-negligible fraction of subjects chose to disclose their past, irrespective of whether disclosing was costly or not.[27] The overall fraction of disclosers in the C-RM treatment was 23.0% (Table 1). As consistent with Hypothesis 2(a), the disclosure rate is significantly higher in the F-RM than in the C-RM treatment (Part II in Appendix Table C.1): the fraction of disclosers was 53.9% in the F-RM treatment (Table 1).[28] The subjects' average beliefs on the percentages of disclosers transited almost parallel to the actual percentages of disclosers in the C-RM and F-RM

[27] A regression analysis shows that subjects' frequency of disclosing in each of the C-RM and F-RM treatments is significantly different from 0. See the estimates of the constant terms in Panel I, Appendix Table C.1.
[28] See Panel II in Appendix Table C.1.

treatments (Panels I(c) and II(c) in Figure 2).[29] The trends of average contributions in the C-RM and F-RM treatments are similar to each other: average contributions gradually declined over time as usual free-riding dynamics seen in the literature (e.g. Chaudhuri, 2011; Ledyard, 1995).

**Result 1**: *(i) Non-negligible fractions of subjects disclosed their past in the C-RM and F-RM treatments. Subjects correctly formed beliefs on the percentages of disclosers among peers. (ii) The average contributions in the C-RM and F-RM treatments both gradually declined over time.*

What were the effects of information revealed by subjects? In order to study the impact of (free or costly) information disclosure on the community's cooperation norms, we conducted an additional treatment, which we call the "Baseline" treatment (Table 1). In the Baseline treatment, subjects were not allowed to disclose their past, were just randomly matched with another subject in a session, and then played the same two-person public goods game in each period (see Eq. (1) for the payoff consequence). The number of periods was 20 as in the other treatments. The rest of the design pieces were identical to the C-RM and F-RM treatments. As shown in Figure 2 [the dashed lines in Panels I(a) and II(a)], average contributions steadily declined from period to period in the BASELINE treatment, similar to the C-RM and F-RM treatments, but the level of cooperation in the former was much lower than that in the latter. The overall average contribution amount was 3.91 ECUs in the Baseline treatment (Table 1). This shows a positive effect of information disclosed in the C-RM and F-RM treatments.

**Result 1:** *(iii) The average contributions declined over time in the BASELINE treatment, similar to the C-RM and F-RM treatments; but the levels of cooperation were always much lower in the former treatment than in the latter two treatments.*

When the sorting device was present, significantly larger fractions of subjects disclosed their last-period contribution amounts to join the reputation community, compared with the corresponding random-matching treatments (again see Figure 2).[30] 52.8% and 70.7% of subjects

[29] This is consistent with the assumption we imposed in analyzing the reputation building behavior of a material payoff maximizer in the presence of conditional cooperators in Section 3.
[30] See Panel I in Appendix Table C.1 for a statistical test.

in the C-Sorting and F-Sorting treatments, respectively, sorted into the reputation community (Table 1). The higher disclosure rate in the C-Sorting treatment, relative to the C-RM treatment, supports Hypothesis 2(b), but the higher disclosure rate in the F-Sorting treatment, relative to the F-RM treatment, is not consistent with Hypothesis 2(c).

The pattern of subjects' contribution behaviors is consistent with Hypothesis 3(b) for the C-Sorting and C-RM treatments, but not for the F-Sorting and F-RM treatments (Figure 2). Our date shows that the average contributions were much higher in the C-Sorting than in the C-RM treatment, but the levels of cooperation were similar between the F-Sorting and F-RM treatments.

**Result 2:** *(i) Significantly larger fractions of subjects disclosed their past in the C-Sorting and F-Sorting treatments, compared with the C-RM and F-RM treatments, respectively. (ii) The average contributions were higher in the C-Sorting than in the C-RM treatment. (iii) However, they were at similar levels between the F-Sorting and F-RM treatments.*

A close look at the data indicates that Result 2(ii) was driven by the high efficiency in the reputation community in the C-Sorting (Panel I(a) in Figure 2). Disclosers in the reputation community continuously contributed more than half of their endowment, ranging from around 10 to 15 ECUs on average, to their joint accounts in the C-Sorting treatment. The high cooperation norms were well sustained, except the end periods (Andreoni, 1988).[31] This pattern is clearly different from that in the C-RM treatment as well as in the anonymous community in the C-Sorting treatment (Panel I(a), Figure 2).[32] Although the disclosers who joined the reputation community had to pay one ECU for disclosure, the average per-period payoff in the C-Sorting treatment (26.01 ECUs) was much higher than that in the C-RM treatment (23.40 ECUs). We note that in the F-Sorting treatment, where the sorting process was free, the average contributions

[31] The trend of average contributions was overall in a declining trend due to the end game effects. However, there is no sign of decline in the reputation community in the C-Sorting treatment if we exclude a few end periods. For instance, the period number has a coefficient of .00043 when we perform a session fixed effect linear regression with standard errors clustered by session, where the period number is the only regressor, using session-average data.

[32] The overall average contribution of the disclosers in the reputation community is 11.96 ECUs, which is much higher than the average contribution in the C-RM treatment, 6.02 ECUs. The average contribution in the anonymous community in the C-Sorting treatment was almost the same as the average contribution in the C-RM treatment (Table 1, Panel I(a) in Figure 2). See Table 2 for a statistical test of the difference.

in the reputation community hovered between 8 ECUs and 11 ECUs during the first ten periods and then between 6 ECUs and 9 ECUs during the second ten periods. As shown in Panel II(a) of Figure 2, the average contributions in the reputation community were also on average in a declining trend, unlike the ones in the C-Sorting treatment. We also note that low cooperation norms were commonly observed in the anonymous community in both of the two sorting treatments (Panels I(a) and II(a) in Figure 2).

**Result 2:** *(iv) Result 2(ii) was driven by strong cooperation norms well sustained in the reputation community in the C-Sorting treatment.*

In Section 4.2, we will formally study the treatment effects of information disclosure and sorting on subjects' contribution behaviors.

*Disclosure and Last Period Contribution Behavior:*

Lastly, we overview the dynamics of subjects' disclosure or sorting decisions in each treatment. The disclosure rate (the percentage of the cases in which subjects chose to disclose [sort into the reputation community] in a random-matching [sorting] treatment) was in a decreasing trend in the C-RM treatment, but it stayed stable in the C-Sorting, F-RM and F-Sorting treatments (Panels I(c) and II(c) in Figure 2). This is perhaps because disclosure involves a cost and no sorting mechanism is present in the C-RM treatment and thus subjects gradually learn that it is not worth one ECU.

Subjects' disclosure decisions in period $t$ were largely affected by their own contribution behaviors in period $t-1$, rather than their matched partners' decisions in period $t-1$. As shown in Appendix Table C.2, the lower amounts disclosers contributed to the joint account in period $t-1$, the more likely they were to become non-disclosers in period $t$ in all the four treatments. Similarly, the higher amounts non-disclosers contributed to the joint account in period $t-1$, the more likely they switched to become disclosers in period $t$ in the C-RM, F-Sorting and F-RM treatments.[33] This suggests that subjects' disclosure or sorting decisions were more motivated by own intentions to cooperate in the future, or hide low contribution histories or escape from exploited partners, rather than passive motives based on partners' action choices taken to the subjects. Some subjects switched back and forth between the two communities in the C-Sorting and F-Sorting treatments (Appendix Figure C.1). On average 10% to 20% of subjects moved from the reputation to the anonymous communities; and similar fractions of subjects moved from the anonymous to the reputation communities in each period.

*4.2. Treatment Effects of Voluntary Information Disclosure and Sorting*

Our approach to study the treatment differences is to estimate a regression model (columns (1) and (2) in Table 2), in which the dependent variable is (session-average) contribution amounts in each period. Independent variables include treatment dummies, and the reference group is contribution amounts in the Baseline treatment. The period number {= 2, 3, …, 20} is also controlled for in column (2). The estimation shows that the impact of information disclosure or sorting depends on the frequency of disclosing. The impact of costly disclosure is not significant in the C-RM treatment, whose disclosure rate was 23.0%, the lowest of the four treatments (see the coefficient estimates of the C-RM dummy). In contrast, in the F-RM, C-Sorting, and F-Sorting treatments, whose disclosure rates were more than 50%, the impact of free disclosure or that of sorting, whether costly or free, is significant. Overall, there were no treatment differences in the level of cooperation between the three treatments.

Considering that some subjects in the C-RM and C-Sorting treatments paid fees for disclosure and sorting, respectively, we also performed the same regressions while setting the dependent variable as (session-average) payoffs in each period. As shown in Appendix Table C.3, we found significant effects of voluntary disclosure in the F-RM treatment and sorting in the C-Sorting and F-Sorting treatments in increasing payoffs (see columns (1) and (2) of Table C.3).

[33] Last-period partners' contribution behaviors did not significantly affect subjects' disclosure behaviors, except for disclosers in the F-Sorting treatment. The disclosers in the F-Sorting treatment were more likely to stay in the reputation community in the following period, the higher amounts their matched partners contributed.

**Result 3:** *(i) Costly disclosure in the C-RM treatment does not either enhance cooperation or increase payoffs significantly. However, (ii) free disclosure in the F-RM treatment and sorting in the C-Sorting and F-Sorting treatments have significant impact on raising people's levels of cooperation and payoffs.*

Result 2(iv) is also confirmed by a regression analysis. As shown in column (3), Table 2, the average contribution is significantly higher in the C-Sorting reputation community than in the C-Sorting anonymous community or in the C-RM treatment. This picture changes when disclosure does not involve a cost. The average contribution in the F-Sorting reputation community is *not* significantly different from that in the F-RM treatment, although the former is significantly higher than that in the F-Sorting anonymous community. The same also holds when subjects' average payoffs, instead of their average contributions, are compared (see column (3) in Appendix Table C.3).

**Result 3**: *(iii) The average contribution is significantly higher in the reputation than in the anonymous community, regardless of whether the sorting is free or costly. (iv) When disclosing information involves a cost, the average contribution in the reputation community is significantly higher than that in the random-matching community of the C-RM treatment.*

We note that the average contribution in the reputation community is significantly higher at the 10% level in the C-Sorting than in the F-Sorting treatment (column (3), Table 2). However the average payoffs in the reputation community are not significantly different between the two treatments (column (3), Table C.3) due to the costs spent by disclosers in the C-Sorting treatment.

*4.3. Subjects' Disclosure Decisions and Action Choices in the C-RM and F-RM treatments*

This subsection is devoted to an analysis of subjects' behaviors in the C-RM and F-RM treatment which may have driven Results 1 to 3.

*Disclosers' High Willingness to Contribute:*

The difference in the frequency of subjects' disclosure (Table 1) and Result 3 seem to suggest that the contribution behaviors of disclosers can be different from those of non-disclosers. In Section 4.3, we will first explore the relationship between subjects' disclosure decisions and their action choices. Figure 3 reports the average contribution amounts by disclosure decision in the C-RM and F-RM treatments. It indicates that disclosers on average contributed 11.0 ECUs and 10.7 ECUs in the C-RM and F-RM treatments, respectively, both of which are significantly higher than the average contribution amounts of non-disclosers in these two treatments (which are 4.5 ECUs and 4.1 ECUs).[34] These suggest that subjects' disclosure decisions and their contribution amounts are closely linked to each other as assumed in the discussions in Section 3.

**Result 4:** *Disclosers are significantly more likely than non-disclosers to contribute large amounts to the joint accounts in the C-RM and F-RM treatments.*

*Cooperation Types and Strategic Reputation Building Behaviors:*

Not only in the F-RM but also in the C-RM treatments, our data shows that some selfish subjects strategically contributed positive amounts to the joint account to build cooperative reputations and then disclosed the contribution amounts in the following periods. The conditional contribution schedule (Fischbacher *et al.*, 2001) elicited from each subject can be used for this analysis.[35] In our detailed analysis below, we focus on two types: "conditional cooperators" and "free riders." Similar to Fischbacher *et al.*, we define those whose own contribution amounts and the others' average contribution amounts are significantly positively correlated at least at the 5% level (according to spearman's $\rho$ correlation coefficients) as the conditional cooperators. We define those whose own contribution amounts are always zero as free riders. Based on these classification criteria, 58.3% and 12.5% of subjects are classified as conditional cooperators and

[34] We conducted a linear regression (with robust standard errors clustered by session) in which the dependent variable is subject $i$'s contribution amount in period $t$ and independent variables includes a dummy which equals 1 if $i$ disclosed and 0 if $i$ did not disclose his or her last-period contribution amount. Individual fixed effects were added to control for the panel structure. We found that the dummies obtain significantly positive coefficients at the 5% and 10% level in the C-RM and F-RM treatments, respectively.

[35] The average conditional contribution schedule of the overall subject population group falls in line with the standard conditional cooperator type: own contribution amounts are significantly positively increasing in the others' average contribution amounts (Appendix Figure C.2).

free riders, respectively, in the C-RM treatment. Likewise, 63.6% and 20.5% of subjects are classified as conditional cooperators and free riders, respectively, in the F-RM treatment.

The percentages of conditional cooperators and free riders who costly disclosed their past in the C-RM treatment are on average 25.8% and 21.1%, respectively. These two percentages in the F-RM treatment are 55.3% and 50.3%, respectively. Thus, a non-negligible fraction of free riders did disclose their last-period action choices, like conditional cooperators.

The fit of the classification method by Fischbacher *et al.* can be examined by looking at subjects' contribution amounts in period 20 (the final period of the experiment). In period 20, a subject would contribute nothing if he or she is purely selfish. Panels 1(b) and 2(b) of Figure 3 report the average contributions in period 20. Two clear features were found, among others. First, disclosers on average contributed more than non-disclosers even when only period 20 is considered.[36] The average contributions of disclosers were 9.2 ECUs and 7.4 ECUs in the C-RM and F-RM treatments, respectively. Second, and as importantly, the average contributions of free riders in period 20 were much lower than those of conditional cooperators. For instance, the average contributions of free riders in the C-RM treatment were 0.0 ECUs when choosing to disclose (0.6 ECUs when not disclosing their last-period actions). The second observation implies that the classified types are good indicators to measure subjects' contribution behavior.

Panels 1(a) and 2(a) of Figure 3 report average contribution amounts across all periods but periods 1 and 20 by disclosure decision. It indicates that not only conditional cooperators but also free riders contributed large amounts in the C-RM and F-RM treatments when they disclosed their last-period contribution amounts. The differences in the average contribution between conditional cooperators and free riders are not significant in both of the treatments (see Wald test results in Appendix Table C.4). This suggests that as discussed in Section 3, some free riders indeed mimicked the behavior of conditional cooperators by strategically contributing large amounts and disclosing them in the C-RM and F-RM treatments.

[36] The ratios of the average contribution by disclosers to that by non-disclosers are around 2.9 and 8.2 in the C-RM and F-RM treatments, respectively (see the "All subjects" bars in the two panels).

**Result 5:** *(a) Regardless of disclosure decision, free riders contributed very little in period 20, compared with conditional cooperators, in the C-RM and F-RM treatments. (b) However, free riders contributed almost similar to conditional cooperators during the course of their plays (periods before period 20) when they disclosed the last-period contribution amounts.*

Why did some conditional cooperators choose not to disclose? One possibility is that they were pessimistic about their peers' contribution behaviors. As shown in Figure C.4, we partially confirm this possibility: conditional cooperators who less frequently disclosed formed lower beliefs on their matched peers' contribution behaviors, compared with the same types who frequently disclosed (Panels a and c). However, the differences in the average belief are not significant in both the C-RM and F-RM treatments. Instead, Figure C.4 indicates that among conditional cooperators, infrequent disclosers contributed significantly less than frequent disclosers, when they selected to disclose, in these two treatments. This seems to suggest that some conditional cooperators attempted to exploit other cooperators, instead of encouraging others to select conditional cooperative strategies. This may mean that there are other aspects in human cooperativeness that are not captured by the classification method by Fischbacher *et al*.

In summary, whether conditional cooperators or free riders, some subjects attempted to build cooperative reputations in the C-RM and F-RM treatments. The difference in the efficiency between the C-RM and F-RM treatments can be interpreted that the cost of disclosure discouraged a large percentage of subjects from disclosing and cooperating with their peers.

### *4.4. Subjects' Sorting Decisions and Action Choices in the C-Sorting and F-Sorting treatments*

As discussed, subjects successfully cooperated with each other in the reputation community, but not in the anonymous community, in the two sorting treatments. What drove the successful cooperation achieved in the reputation community? With a closer look, we found that although the level of cooperation was higher in the reputation than in the anonymous community in the F-Sorting treatment, the cooperation norms in the reputation community was lower, compared with the C-Sorting reputation community. Why did the performance of the reputation community differ between the two treatments? This section will study these questions.

*Composition of Cooperation Types by the Community:*

The C-Sorting and F-Sorting treatments have sorting mechanisms. One possible answer to the above two questions is that the composition of cooperation types are different between the reputation and anonymous communities.[37] However, our data does not support this hypothesis. The percentages of conditional cooperators and free riders are not significantly different between the reputation and anonymous communities in both the C-Sorting and F-Sorting treatments (Appendix Figure C.3). Moreover, as similar to Result 5(b), free riders contributed amounts similar to conditional cooperators in the reputation community.[38] Some free riders' reputation building behavior is not surprising, particularly because they were assured to be paired with other disclosers. These analyses suggest that the stronger cooperation norms emerged in the reputation community, especially in the C-Sorting treatment, were not due to the self-selection of cooperation-oriented subjects.

**Result 6:** *The percentages of conditional cooperators and free riders in the reputation community are not significantly different from those in the anonymous community in the C-Sorting and F-Sorting treatments. Moreover, the contribution behaviors of conditional cooperators are not significantly different from those of free riders in the sorting treatments.*

*Subjects' Action Choices and Beliefs on their Peers' Behaviors:*

The next possible factor that was behind the high efficiency in the reputation community is subjects' higher beliefs on their matched disclosers' contribution amounts.

First, our data shows that subjects' beliefs on their matched partners' contribution amounts significantly affected their own contribution behavior. This is true for both the two sorting treatments and the two random-matching treatments (Table 3). Panels I(b) and II(b) in

[37] It has been demonstrated that subjects cooperate significantly more if they are grouped with like-minded cooperative types in dilemma games (e.g., Gunnthorsdottir *et al.*, 2007; Gächter and Thöni, 2010).
[38] Conditional cooperators and free riders on average contributed 11.4 ECUs and 10.2 ECUs, respectively, to the joint accounts in the C-Sorting reputation community. Conditional cooperators and free riders on average contributed 9.3 ECUs and 7.5 ECUs, respectively, to the joint accounts in the F-Sorting reputation community.

Figure 2 report the trends of the subjects' beliefs in the C-Sorting and F-Sorting treatments. The data revealed that disclosers' contribution amounts fluctuated parallel to their beliefs on matched partners' contributions (see the red lines with diamonds in Panels (a) and (b)).[39] The similar holds for non-disclosers (see the blue line with triangles in Panels (a) and (b)).[40] A regression analysis confirms that contribution amounts and subjects' beliefs are significantly positively correlated for each of disclosers and non-disclosers in both the C-Sorting and F-Sorting treatments (columns (3) to (6) in Table 3(a)). The highly positive correlations between own actions and the beliefs resonates with the idea that people experience psychological satisfaction from mutual cooperation (e.g., Fehr and Gächter, 2000, 2002; Charness and Rabin, 2002; Falk *et al.*, 2005; Kamei and Putterman, forthcoming; Rilling *et al.*, 2002; Decety *et al.*, 2004). The impact that subjects' beliefs have on their own contribution behaviors is similarly observed in the C-RM and F-RM treatments (columns (1) and (2) in Table 3(a)).

Second, our data shows that subjects formed beliefs based on their matched partners' last-period contribution amounts when the past contribution amounts were revealed to them. Table 3(b) reports a regression analysis in which the dependent variable is subject $i$'s belief on her period $t$ partner's contribution amount and independent variables include the partner's reputation score. First, in the C-RM and F-RM treatments, subjects formed significantly higher beliefs when they were matched with disclosers than when they were matched with non-disclosers. In addition, when subjects were matched with disclosers, their beliefs were positively correlated with the disclosers' last-period contribution amounts (Panel (b1)). Second, likewise, in the reputation communities of the two sorting treatments, significantly positive correlations between disclosers' beliefs and period $t$ partners' last-period contribution amounts were observed (Panel (b2)). Panel I(b) of Figure 2, for example, shows that the much higher beliefs formed by

[39] The disclosers' average belief in the C-Sorting treatment was 12.15 ECUs, which was only .2 ECUs higher than the average of their own contribution amounts. Likewise, the disclosers' average belief in the F-Sorting treatment was 8.75 ECUs, which was only .28 ECUs lower than the average of their own contribution amount.

[40] For instance, the non-disclosers in the C-Sorting treatment on average believed that their partners would contribute 7.10 ECUs, which was only a little higher than the non-disclosers' average contribution, but it is not significantly different from the actual contribution amount.

disclosers in the C-Sorting treatment are almost parallel to the high contribution amounts their matched partners made in the last period which were informed to disclosers.

These analyses suggest that the success of high efficiency in the reputation community may lie on the sorting mechanism where disclosers are matched with other disclosers, because the sorting mechanism helps subjects keep high beliefs on their partners' contribution behaviors.

**Result 7:** *(a) Subjects' contribution amounts and beliefs on their partners were significantly positively correlated in all treatments. (b) Disclosers' beliefs and their partners' last-period contribution amounts were significantly positively correlated in the C-Sorting and F-Sorting reputation communities. (c) Subjects formed significantly higher beliefs when they were matched with disclosers than otherwise; and the beliefs on the disclosers' contribution amounts were significantly positively correlated with the partners' last-period contribution amounts in the C-RM and F-RM treatments.*

*Performance Differences in the Reputation Community between the Two Sorting Treatments:*

As discussed earlier, Panels I(a) and II(a) of Figure 2 indicate a difference in the evolution of cooperation in the reputation community between the C-Sorting and F-Sorting treatments. These figures and Result 7 suggest that the difference in the level of cooperation between the two treatments could be caused by the difference in subjects' beliefs on the partners' cooperativeness, caused by the difference in the partners' last-period contribution amounts. A likely reason for this difference is the free entrance to the reputation community in the F-Sorting treatment. In the literature of endogenous group formation, as discussed in Section 3, it is known that selfish individuals may chase cooperative ones attempting to exploit them and as a result cooperation may easily collapse (Ahn *et al.*, 2008, 2009; Ehrhart and Keser, 1999). The lack of the cost for sorting in the F-Sorting treatment may strengthen this negative effect of mobility.

Appendix Figure C.5 reports subject-by-subject average contribution amounts in the reputation community in the C-Sorting and F-Sorting treatments. This shows that most subjects frequently switched back and forth between the reputation and anonymous communities. However, more subjects in the F-Sorting treatments frequently joined the reputation community, compared with the C-Sorting treatment. Appendix Figure C.6 reports the history of each subject's sorting decisions. It reveals that unlike the C-Sorting treatment, most subjects in the F-Sorting treatment switched back and forth between the two communities with different cycles.[41] Despite that on average around 25% to 35% of subjects moved between the reputation and anonymous communities in each period for both the two sorting treatments (see Appendix Figure C.1 and the discussion in Section 4.1), it seems that only a small set of subjects in the C-Sorting treatment switched back and forth between the reputation and anonymous communities.

The percentages of subjects who joined the reputation community more than or equal to ten times (except period 20), which we call the "frequent disclosers" hereafter, are 52.3% and 77.3% in the C-Sorting and F-Sorting treatments, respectively. The two percentages are significantly different (Two-sample test of proportions, two-sided $p$-value = .0141). Appendix Figure C.5 also indicates that a larger percentage of the frequent disclosers contributed large amounts to the joint account in the C-Sorting treatment, compared with the F-Sorting treatment. For example, the percentages of the frequent disclosers that contributed amounts less than 10 ECUs are 11.4% and 43.2% in the C-Sorting and F-Sorting treatments, respectively. The small percentage in the C-Sorting treatment is striking. The two percentages are significantly different (Two-sample test of proportions, two-sided $p$-value = .0107). These analyses suggest that a significantly larger fraction of subjects had intentions to exploit high contributors and joined the reputation community when joining is free than when it is costly, and they sometimes escaped into the anonymous community after engaging in such exploitable behaviors.

**Result 8:** *A significantly larger percentage of those who attempt to exploit high contributors joined the reputation community in the F-Sorting than in the C-Sorting treatment.*

Result 8 implies that the reputation community can be more unstable in the F-Sorting than in the C-Sorting treatment because of the discriminating strategies [Results 7(a) and (b)];

[41] There is also quantitative evidence to support this observation. First, disclosers in period $t-1$ in the C-Sorting treatment were more likely to stay as a discloser in period $t$, compared with those in the F-Sorting treatment (Remark in Appendix Figure C.6). Second, non-disclosers in the C-Sorting treatment did not switch to a discloser in the following period when they contributed large amounts to the joint account (columns (a3) in Appendix Table C.2).

and some subjects' low contribution events would spread to other subjects. This contagion process can explain the difference in the cooperation dynamics in the reputation community between the C-Sorting and F-Sorting treatments.

Lastly, we note that although the efficiency was significantly higher in the reputation than in the anonymous community in the C-Sorting treatment, a non-negligible fraction of the subjects stayed away from the reputation community. This phenomenon cannot be explained by cooperation types (Result 6). What may account for this phenomenon? The data shows that the reputation community in the C-Sorting treatment, unlike the F-Sorting treatment, may have attracted subjects with more optimistic expectation about their peers' action choices (see Appendix Table C.6). Specifically, the frequent members of the reputation community had higher expectations about their peers' contribution amounts than those who less frequently joined did in the same reputation community.[42,43] This is suggestive evidence only, however, since the effect is no longer significant once session clustering is added in Table C.6.

*4.5. Little Reliance on the Contagious Punishment Strategy*

Our results in Sections 4.3 and 4.4 are consistent with the idea that some subjects are conditional cooperators and selfish types mimic the behaviors of conditional cooperators. We discussed in Section 3 that cooperation could evolve also with the contagious punishment strategy (grim trigger strategy). As will be briefly reported in this subsection, a close look at each subject's history of action choices suggests that no subjects followed such a strict strategy.

If a subject follows the contagious strategy as described in Section 3, the subject would become a 0-contributor from period $t$ if (a) her partner $j$ in period $k$, where $k < t$, contributed 0 in that period, (b) the partner disclosed but had a reputation score of 0 (i.e., $S_{j,k} = 0$), or (c) her partner was a non-discloser. In our experiment, the likelihood that a subject encountered a non-

[42] In the regression analysis, we used each subject's session-average beliefs on their matched peers' contribution amounts in the reputation community or anonymous community as the dependent variable. The independent variable includes the total number of periods in which the subject joined the reputation community before period 20.
[43] Also see Panel (b) of Appendix Figure C.4. These may mean that subjects' beliefs on their peers' behaviors determined subjects' sorting decisions. However, we acknowledge that the direction of causality may be opposite: subjects' beliefs on the high cooperation norms in the reputation community were mainly formed by their experiences in the reputation community (Result 7).

discloser or a discloser with a low reputation score, or the subject was defected by zero contributions from her partner was large. A calculation shows that if subjects followed the contagious strategy, all subjects would have become 0-contributors from periods 5, 9, 9 and 6 in the C-RM, F-RM, C-Sorting and F-Sorting treatments, respectively. This was clearly not the case in the experiment (see Figure 2). We also checked how many subjects in each treatment turned to complete 0-contributors after experiencing at least one of the three events above, (a), (b) and (c). In this calculation, we excluded subjects who acted on the "always to defect" strategy during the entire experiment.[44] We found that no subjects in each of the four treatments became complete 0-contributors after experiencing such events the first time.

*4.6. Cost of Soring or Information?*

The performance in the reputation community was clearly different from that in the anonymous community in the C-Sorting treatment. It also performed much better than in the C-RM treatment. As a result, the C-Sorting treatment has significantly higher efficiency than the C-RM treatment (Result 2(ii)). This can be due to two effects: (A) the impact of information and (B) the impact of the *cost* to be matched with like-minded others. If the presence of a sorting cost is one key factor for the superior performance of the C-Sorting treatment, then a highly efficient community may emerge and subjects may enjoy high cooperation norms in that community, even if we eliminate the element of the reputation mechanism from the C-Sorting treatment. However, there is also a possibility that cooperation may not evolve if reputational information is removed, because then subjects cannot distinguish between high and low contributors. In order to divide the impact of costly sorting into impact (A) and impact (B), we conducted an additional treatment, called "C-Sorting-N" (Costly Sorting, No Disclosure). In this additional treatment, each subject who paid one ECU was matched with another who paid one ECU. However, their last-period contribution amounts were *not* revealed (see Appendix E for the instructions). We

[44] The number of subjects who acted on the "always to defect" strategy was much less in the experiment. There were four, two, zero, and one subject(s) in the C-RM, F-RM, C-Sorting, and F-Sorting treatments, respectively.

call the community of those who paid (did not pay) the fee the "paid community" ("non-paid community") in the paper. The rest of the design was identical to the C-Sorting treatment.

On average 53.6% of subjects, whose percentage is almost the same as that in the C-Sorting treatment, paid the fee to join the paid community (Table 1). As in the C-Sorting treatment, the average contributions were clearly different between the paid community and the non-paid community (Table 1, Figure 4). The difference in the level of cooperation is significant and huge, around 7 ECUs (Table 1, columns (3) and (4) in Appendix Table C.7(I)). Nevertheless, the average contribution is 2.51 ECUs lower in the paid community than in the C-Sorting reputation community (see again Table 1). As for the treatment differences, the average contribution in the C-Sorting-N treatment is 6.31 ECUs, around 61% higher than that in the Baseline treatment.[45] However, the former is significantly lower than that in the C-Sorting treatment (see columns (1) and (2) in Panel I, Appendix Table C.7). Qualitatively the same treatment difference is also found when average payoffs are compared between the C-Sorting and C-Sorting-N treatments (see columns (1) and (2) in Panel II, Appendix Table C.7).

In summary, these additional analyses suggest that there is positive impact of a cost to be matched with like-minded others; however, the presence of the reputation mechanism is crucial for the performance in the C-Sorting treatment.

**Result 9:** *(i) In the C-Sorting-N treatment, not only average contributions but also average payoffs were significantly higher in the paid than in the non-paid community. However, (ii) as a whole, the average contribution and average payoff were both significantly lower in the C-Sorting-N than in the C-Sorting treatment.*

### *4.7. Significance of Reputation Mechanisms*

We found that the distributions of cooperation types were not different between the reputation and anonymous communities (Section 4.4). We discussed that the free entrance to the reputation community may have encouraged some subjects to exploit cooperative subjects in the

[45] The difference is significant, according to a regression analysis, only when the Period Number variable is controlled for (see columns (1) and (2) in Panel I, Appendix Table C.7).

reputation community in the F-Sorting treatment. But what happens if there is *no* anonymous community, subjects are *not* allowed to switch between communities and *are always forced* to disclose their past in the interactions? If the cost of mobility between the two communities and the reputation mechanism are the key reasons for the high efficiency in the reputation community in the C-Sorting treatment, subjects could achieve a high level of cooperation as a whole once we eliminate such mobility, under the condition in which the reputation mechanism is present. In this section, as a last analysis, we report the results from an additional treatment to check the impact of reputation mechanisms without mobility. We call this additional treatment the "Info" ("Information Automatic Disclosure") treatment; and we compare it with the C-Sorting treatment. At the onset of the additional sessions, subjects were randomly assigned to a community whose size was six (the community size is on average six also in the reputation community of the C-Sorting treatment) and the community composition was kept the same during the entire experiment. Two sessions, each with 12 subjects, were conducted so that we can obtain four independent community data as in the C-Sorting reputation community. The number of interactions was 20. In each period, subjects were randomly matched with another in their own community and played the same public goods game (Eq. (1)). From period 2, each subject's last period contribution amount was always automatically revealed to their own partner.

As shown in Figure 5, the trend of average contributions in the Info treatment was very similar to that in the C-Sorting reputation community. The average contribution across all periods was 11.52 ECUs, which is almost the same as that in the C-Sorting reputation community – 11.96 ECUs (Table 1). This confirms our earlier discussion that the easiness of mobility and the presence of an anonymous community may lead to a decline of cooperation norms. Nevertheless, the average contribution in the C-Sorting anonymous communities is significantly lower, compared with the Info treatment (column (3) in Appendix Table C.8(I)). As a result, the difference in average contribution between the C-Sorting and Info treatments was more than 2 ECUs, although the treatment difference is not significant (columns (1) and (2) in Table C.8(I)). Moreover, the average payoff is significantly higher in the Info than in the C-

Sorting treatment at the 10% level due to some subjects' cost spent in joining the reputation community and the presence of the inefficiency anonymous community (columns (1) and (2) in Table C.8(II)).

**Result 10:** *(i) The average contribution in the Info treatment is at almost similar levels to that in the C-Sorting reputation community, but the former is significantly higher than that in the C-Sorting anonymous treatment. Due to the presence of an inefficient anonymous community and the cost spent in sorting, (ii) the average payoff is weakly significantly higher in the Info than in the C-Sorting treatment.*

Result 10 underscores negative aspects of people's endogenous sorting and voluntary information disclosure and potentially a desire to have some mechanisms to overcome the observed inefficiency in a community.

## 5. Conclusions

This study let subjects play a finitely repeated two-person public goods game where each subject is given an option to disclose their past behavior. Our experiment showed that voluntary information disclosure is not enough to enhance cooperation if disclosure involves a cost and a sorting mechanism is not present. However, it is indeed helpful for the evolution of cooperation if the disclosure is made for free or if subjects are able to join a community with the reputation mechanism through information disclosure. A close look at the data revealed that the failure of cooperation with costly information disclosure was caused by sufficiently low disclosure rates and the non-disclosers' opportunistic behaviors, among others. When subjects were given an opportunity to sort into the reputation community, a stable size of reputation community emerged and those who joined the reputation community cooperated with each other at high levels. Nevertheless, the high level of cooperation in the reputation community diminished if sorting into that community was cost-free because those with malicious intentions to exploit high cooperators were more likely to join the reputation community, compared with when sorting was costly. Moreover, we found that the presence of the reputation mechanism was crucial for such positive impact of sorting, but mobility to an anonymous community encouraged some subjects to behave opportunistically although they could have cooperated with each other if they had been forced to always disclose.

We remark that the emergence of an anonymous community in the sorting treatments fits observations in our real life as we see communities (e.g., online platforms) both with and without reputation mechanisms. Our detailed analysis suggests that people's endogenous choices generate these two kinds of communities, although subjects can cooperate with each other if they are forced to always act with reputation mechanisms. What kind of additional institutions on top of the reputation mechanism and sorting would be helpful in reversing such subjects from the anonymous environment to the reputation environment and encouraging them to cooperate with each other? This question would be an interesting direction for future research.

As mentioned earlier, our paper has a useful contribution in the large body of literature on rational cooperation in finitely repeated dilemmas, on reputation mechanisms, and on endogenous group formation. We note that our paper also contributes to the literature on voluntary information disclosure which shows that disclosing private information of products or firms may raise the valuation of them. For instance, Lewis (2011) shows that on eBay motors, there is positive impact of voluntary information disclosure (e.g., photos) on the prices of used cars in auctions. Our experimental setup uses a simultaneous public goods game and it can describe a real-world exchange in which the interaction is taking place while both sides do not know the exact value they will get from the other. This kind of interaction has been increasingly popular in our life (e.g., the sharing economy such as Uber). This simultaneous setup is different from the above marketplace example where sellers move first and buyers move next. Our data implies that voluntary information disclosure or sorting also helps lead to socially beneficial relationships in some simultaneous-move transactions.

## REFERENCES

Ahn, T.K., Mark Isaac, and Timothy Salmon, 2008. "Endogenous Group Formation." *Journal of Public Economic Theory* 10: 170-194.

Ahn, T.K., Mark Isaac, and Timothy Salmon, 2009. “Coming and Going: Experiments on Endogenous Group Sizes for Excludable Public Goods.” *Journal of Public Economics* 93: 336-351.

Aimone, Jason, Laurence Iannaccone, Michael Makowsky, and Jared Rubin, 2013. “Endogenous Group Formation via Unproductive Costs,” *Review of Economic Studies* 80: 1215-1236.

Andreoni, James, 1988. “Why free ride? Strategies and learning in public goods experiments.” *Journal of Public Economics* 37: 291-304.

Andreoni, James, and John Miller, 1993. “Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence.” *Economic Journal* 103: 570-585.

Axelrod, Robert, 1984, *The evolution of cooperation*. New York: BasicBooks.

Axelrod, Robert, and Hamilton William, 1981. “The Evolution of Cooperation.” *Science* 211: 1390-1396.

Bayer, Ralph, 2011. “Cooperation in Partnerships: The Role of Breakups and Reputation.” The University of Adelaide School of Economics Research Paper No. 2011-22.

Bernard, Mark, Jack Fanning, and Sevgi Yuksel, 2017. “Finding Cooperators: Sorting Through Repeated Interaction,” working paper.

Bolton Gary, Elena Katok, and Axel Ockenfels, 2004. “How Effective are Online Reputation Mechanisms? An Experimental Study.” *Management Science* 50: 1587-1602.

Bolton Gary, Elena Katok, and Axel Ockenfels, 2005. “Cooperation among strangers with limited information about reputation.” *Journal of Public Economics* 89: 1457-1468.

Camera, Gabriele, and Marco Casari, 2009. “Cooperation among Strangers under the Shadow of the Future.” *American Economic Review* 99: 979-1005.

Charness, Gary, and Matthew Rabin, 2002. “Understanding Social Preferences with Simple Tests.” *Quarterly Journal of Economics* 117: 817-869.

Charness, Gary, and Yang Chun-Lei, 2014. “Starting Small toward Voluntary Formation of Efficient Large Groups in Public Goods Provision.” *Journal of Economic Behavior & Organization* 102: 119-132.

Chaudhuri, Ananish, 2011. “Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature.” *Experimental Economics* 14: 47-83.

Cooper, Russell, Douglas DeJong, Robert Forsythe, and Thomas Ross, 1992. "Communication in Coordination Games." *Quarterly Journal of Economics* 107: 739-771.

Cox, James, Daniel Friedman, and Steven Gjerstad, 2007. "A tractable model of reciprocity and fairness." *Games and Economic Behavior* 59: 17-45.

Dal Bó, Ernesto, and Pedro Dal Bó, 2014. "“Do the right thing:” The effects of moral suasion on cooperation." *Journal of Public Economics* 117: 28-38.

Dal Bó, Pedro, Andrew Foster and Louis Putterman, 2010. "Institutions and Behavior: Experimental Evidence on the Effects of Democracy." *American Economic Review* 100: 2205-2229.

Decety, Jean, Philip Jackson, Jessica Sommerville, Thierry Chaminade, and Andrew Meltzoff, 2004. "The neural bases of cooperation and competition: an fMRI investigation." *NeuroImage* 23: 744-751.

Devetag, Giovanna, and Andreas Ortmann, 2007. "When and why? A critical survey on coordination failure in the laboratory." *Experimental Economics* 10(3): 331-344.

Duffy, John, and Nick Feltovich, 2002. "Do Actions Speak Louder Than Words? Observation vs. Cheap Talk as Coordination Devices." *Games and Economic Behavior* 39: 1-27.

Duffy, John, and Nick Feltovich, 2006. "Words, Deeds, and Lies: Strategic Behaviour in Games with Multiple Signals," *Review of Economic Studies* 73: 669-688.

Dufwenberg, Martin, and Georg Kirchsteiger, 2004. "A theory of sequential reciprocity." *Games and Economic Behavior* 47: 268-298.

Ehrhart, Karl-Martin, and Claudia Keser, 1999. "Mobility and Cooperation: on the Run." CIRANO working paper 99s-24.

Embrey, Matthew, Guillaume Fréchette, and Sevgi Yuksel, forthcoming. "Cooperation in the Finitely Repeated Prisoner's Dilemma." *Quarterly Journal of Economics*.

Engelmann, Dirk, and Urs Fischbacher, 2009. "Indirect Reciprocity and Strategic Reputation Building in an Experimental Helping Game." *Games and Economic Behavior* 67: 399-407.

Falk, Armin, Ernst Fehr, and Urs Fischbacher, 2005. "Driving Forces of Informal Sanctions." *Econometrica* 73: 2017-2030.

Fehr, Ernst, and Simon Gächter, 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90: 980-994.

Fehr, Ernst, and Simon Gächter, 2002. "Altruistic punishment in Humans." *Nature* 415: 137-140.

Fehr, Ernst, and Klaus Schmidt, 1999. "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics* 114: 817-868.

Fehr, Ernst, and Klaus Schmidt. 2006. ''The Economics of Fairness, Reciprocity and Altruism—Experimental Evidence and New Theories.'' In *Handbook of the Economics of Giving, Altruism and Reciprocity*, edited by S.-G. Kolm and J. M. Ythier, pp. 615-91. North Holland.

Fehr, Ernst, and Tony Williams, 2013. "Endogenous emergence of institutions to sustain," mimeo.

Fischbacher, Urs, 2007. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics* 10: 171-178.

Fischbacher, Urs, and Simon Gächter, 2010. "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments." *American Economic Review* 100: 541-56.

Fischbacher, Urs, Simon Gächter, and Ernest Fehr, 2001. "Are people conditionally cooperative? Evidence from a public goods experiment." *Economics Letters* 71: 397-404.

Gächter, Simon, and Elke Renner, 2010. "The effects of (incentivized) belief elicitation in public goods experiments." *Experimental Economics* 13: 364-377.

Gächter, Simon, and Christian Thöni, 2010. "Social Learning and Voluntary Cooperation among Like-Minded People." *Journal of European Economic Association* 3: 303-314.

Gallo, Edoardo, and Chang Yan, 2015. "The Effects of Reputational and Social Knowledge on Cooperation." *Proceedings of the National Academy of Science USA* 112: 3647-3652.

Greiner, Ben, 2015. "Subject pool recruitment procedures: organizing experiments with ORSEE." *Journal of Economic Science Association* 1: 114-125.

Grimm, Veronika, and Friederike Mengel, 2009. "Cooperation in Viscous Populations – Experimental Evidence." *Games and Economic Behavior* 66: 202-220.

Gunnthorsdottir, Anna, Daniel Houser, and Kevin McCabe, 2007. "Disposition, history and contributions in public goods experiments." *Journal of Economic Behavior & Organization* 62: 304-315.

Gürerk, Özgür, Bernd Irlenbusch, and Bettina Rockenbach, 2006. "The Competitive Advantage of Sanctioning Institutions." *Science* 312: 108-110.

Hatton, Celia, 2015. "China 'social credit': Beijing sets up huge system." *BBC News*, October 26.

Kamei, Kenju, 2017. “Endogenous Reputation Formation under the Shadow of the Future,” *Journal of Economic Behavior & Organization*142: 189-204.

Kamei, Kenju, 2016a. “Joint Decision-Making and Strategic Reputation Building in a Finitely-Repeated Dilemma.” working paper.

Kamei, Kenju, 2016b. “Cooperation and Endogenous Repetition in an Infinitely Repeated Social Dilemma: Experimental Evidence.” working paper.

Kamei, Kenju, and Louis Putterman, forthcoming. “Reputation Transmission without Benefit to the Reporter: a Behavioral Underpinning of Markets in Experimental Focus.” *Economic Inquiry*.

Kamei, Kenju, and Louis Putterman, 2017. “Play it Again: Partner Choice, Reputation Building and Learning from Finitely-Repeated Dilemma Games.” *Economic Journal* 127(602): 1069–1095.

Kandori, Michihiro, 1992. “Social Norms and Community Enforcement.” *Review of Economic Studies* 59(1): 63-80.

Kurzban, Robert, and Daniel Houser, 2005. “Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations.” *Proceedings of the National Academy of Sciences* 102: 1803-1807.

Kreps, David, Paul Milgrom, John Roberts, and Robert Wilson, 1982. “Rational Cooperation in the Finitely Repeated Prisoners’ Dilemma.” *Journal of Economic Theory* 27: 245-252.

Ledyard, John, 1995. “Public goods: A survey of experimental research.” In J. H. Kagel and A.E. Roth (eds.), *The Handbook of Experimental Economics* 111-194, Princeton University Press.

Lewis, Gregory. 2011. “Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors.” *American Economic Review* 101: 1535-1546.

Milinski, Manfred, Dirk Semmann, and Hans-Jürgen Krambeck, 2002. “Reputation helps solve the ‘tragedy of the commons’.” *Nature* 415: 424-426.

Nicklisch, Andreas, Kristoffel Grechenig, and Christian Thöni, 2015. “Information-sensitive Leviathans – the Emergence of Centralized Punishment.” WiSo-HH Working Paper No. 13.

Nowak, Martin, Karen Page, and Karl Sigmund, 2000. “Fairness versus reason in the ultimatum game.” *Science* 289: 1773-1775.

Nowak, Martin, and Karl Sigmund, 1998a. “Evolution of Indirect Reciprocity by Image Scoring.” *Nature* 393: 573-577.

Nowak, Martin, and Karl Sigmund, 1998b. “The dynamics of indirect reciprocity.” *Journal of Theoretical Biology* 194: 561-574.

Page, Talbot, Louis Putterman, and Bulent Unel, 2005. “Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry and Efficiency.” *Economic Journal* 115: 1032-1053.

Rabin, Matthew, 1993. “Incorporating Fairness Into Game Theory and Economics.” *American Economic Review* 83: 1281-1302.

Reuben, Ernesto, and Sigrid Suetens, 2012. “Revisiting Strategic versus Non-strategic Cooperation.” *Experimental Economics* 15: 24-43.

Rilling, James, David Gutman, Thorsten Zeh, Giuseppe Pagnoni, Gregory Berns, and Clinton Kilts, 2002. “A Neural Basis For Social Cooperation.” *Neuron* 35: 395-405.

Rockenbach, Bettina, and Manfred Milinski, 2006. “The efficient interaction of indirect reciprocity and costly punishment.” *Nature* 444: 718-723.

Seinen, Ingrid, and Arthur Schram, 2006. “Social status and group norms: Indirect reciprocity in a helping experiment.” *European Economic Review* 50: 581-602.

Selten, Reinhard, and Rolf Stoecker, 1986. “End Behaviour in sequences of finite prisoner‘s dilemma supergames: a learning theory approach.” *Journal of Economic Behavior and Organization* 7: 47-70.

Sethi, Suresh, and Gerald Thompson, 2006. *Optimal Control Theory* (2nd edition). Springer.

Sobel, Joel, 2005. “Interdependent Preferences and Reciprocity.” *Journal of Economic Literature* 43: 392-436.

Tyran, Jean-Robert, and Lars Feld, 2006. “Achieving Compliance when Legal Sanctions are Non-Deterrent.” *Scandinavian Journal of Economics* 108: 135-56.

Wedekind, Claus, and Victoria Braithwaite, 2002. “The Long-Term Benefits of Human Generosity in Indirect Reciprocity.” *Current Biology* 12: 1012-15.

Wedekind, Claus, and Manfred Milinski, 1996. “Human Cooperation in the Simultaneous and the Alternating Prisoner’s Dilemma: Pavlov Versus Generous Tit-For-Tat.” *Proceedings of the National Academy of Sciences USA* 93: 2686- 2689.

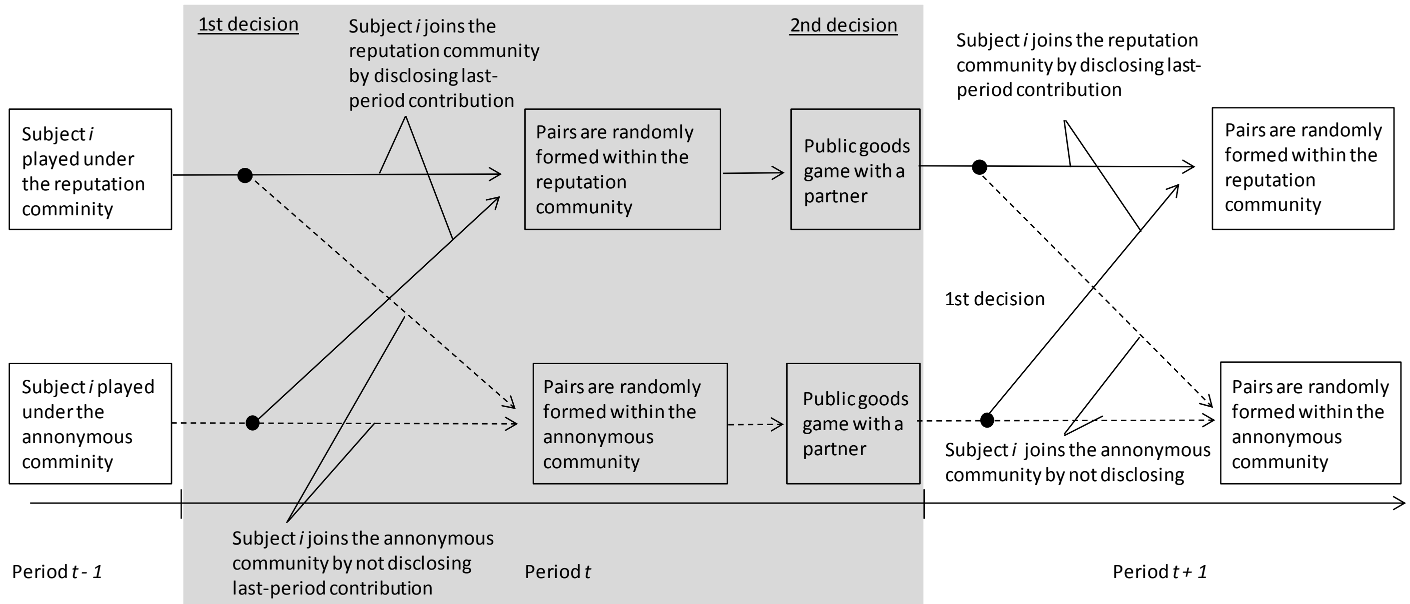
Wedekind, Claus, and Manfred Milinski, 2000. “Cooperation Through Image Scoring in Humans.” *Science* 288: 850-852.

**Table 1.** *Summary of Treatments*

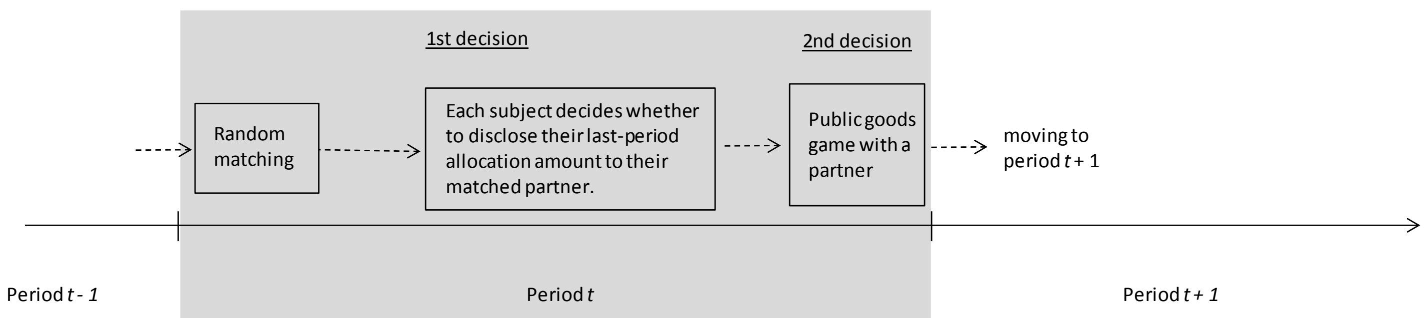
| Treatment name | Costs for disclosing or sorting | Matching Protocol | The number of sessions (subjects) | Fraction of disclosers or those who joined the reputation community | Average contribution | | |
|---|---|---|---|---|---|---|---|
| | | | | | All data | reputation/paid community | anonymous/non-paid community |
| [Main treatments:] | | | | | | | |
| C-Sorting | 1 ECU | Sorting: Each discloser (non-discloser) is matched with another discloser (non-discloser). | 4 (44) | 52.8% | 9.28 (46.4%) | 11.96 (59.8%) | 6.15 (30.7%) |
| C-RM | 1 ECU | Random Matching: Each subject is randomly matched with another subject. | 4 (48) | 23.0% | 6.02 (30.1%) | ---- | ---- |
| F-Sorting | 0 ECUs | Sorting: Each discloser (non-discloser) is matched with another discloser (non-discloser). | 4 (44) | 70.7 % | 7.89 (39.4%) | 8.78 (43.9%) | 5.50 (27.5%) |
| F-RM | 0 ECUs | Random Matching: Each subject is randomly matched with another subject. | 4 (44) | 53.9% | 7.62 (38.1%) | ---- | ---- |
| [Additional treatments:] | | | | | | | |
| Baseline | n.a. | Random Matching: Each subject is randomly matched with another subject. | 4 (44) | n.a. | 3.91 (19.6%) | ---- | 3.91 (19.6%) |
| C-Sorting-N | 1 ECU | Sorting: Each person who paid (did not pay) is matched with another that paid (did not pay). | 4 (44) | 53.6% | 6.31 (31.5%) | 9.45 (47.1%) | 2.35 (11.7%) |
| Info | n.a. | Random Matching: Each subject is randomly matched with another subject. | 2 (24) | n.a. | 11.52 (57.6%) | 11.52 (57.6%) | ---- |
| Total | | | 26 (292) | | | | |

*Notes:* C-Sorting = Costly Sorting. F-Sorting = Free Sorting. C-RM = Costly Disclosure, Random Matching. F-RM = Free Disclosure, Random Matching. C-Sorting-N = Costly Soring, No Disclosure. Info = Information Automatic Disclosure.

**Figure 1.** *Schematic Diagram*



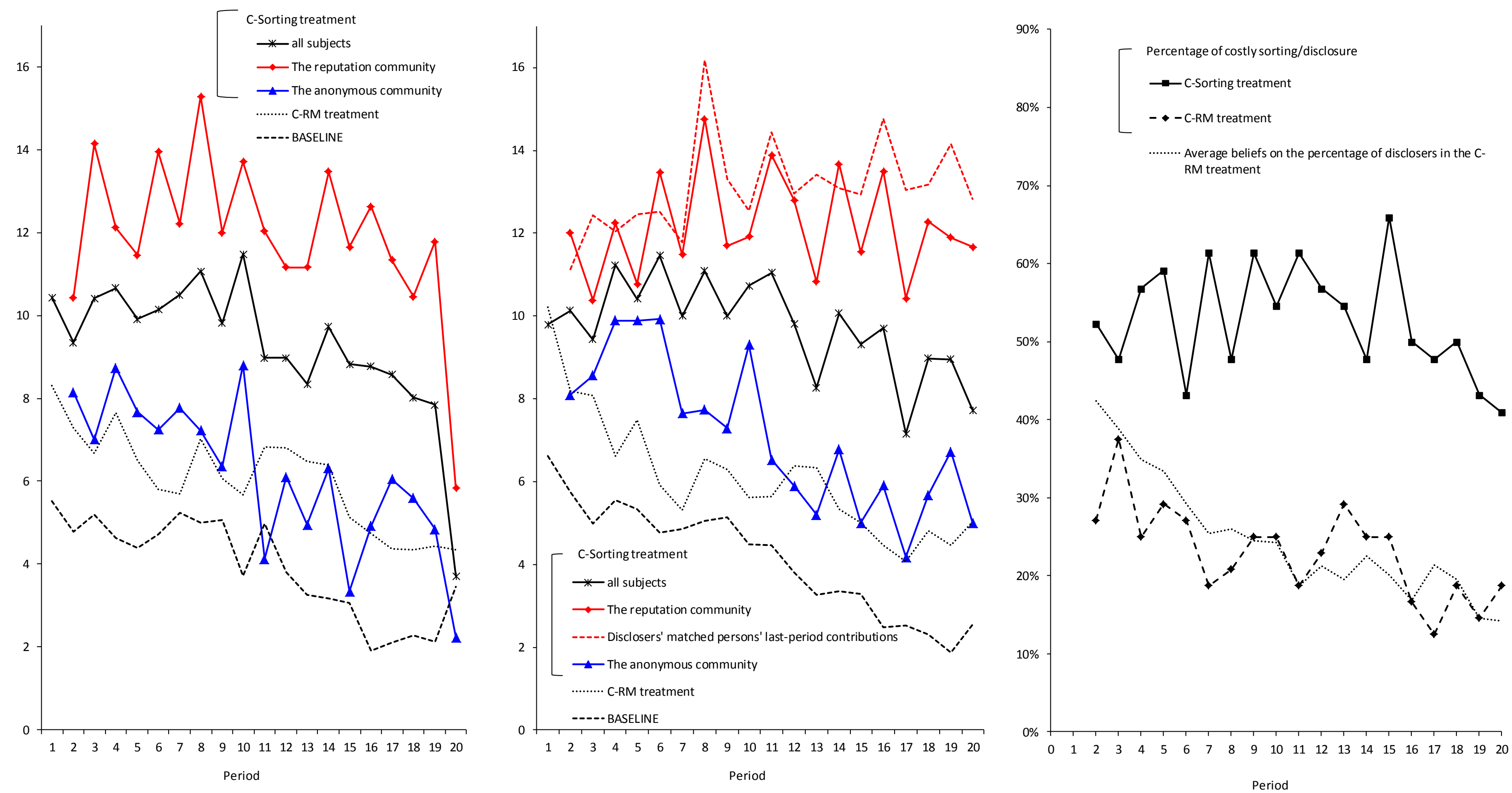

(a) C-Sorting and F-Sorting treatments




(b) C-RM and F-RM treatments

*Notes*: At the onset of the experiment, subjects go through the task to elicit conditional cooperation types. There is no disclosure decision stage in period 1. In the allocation stage, subjects in all treatments are asked to state beliefs on their matched partner's contribution amount in a given period. In the C-RM and F-RM treatments, subjects are also asked to answer guess on how many persons disclosed immediately after their disclosure decision in a given period.

**Figure 2.** *Period-by-Period Average Contributions and the Percentage of the Subjects Who Joined the Reputation Community*

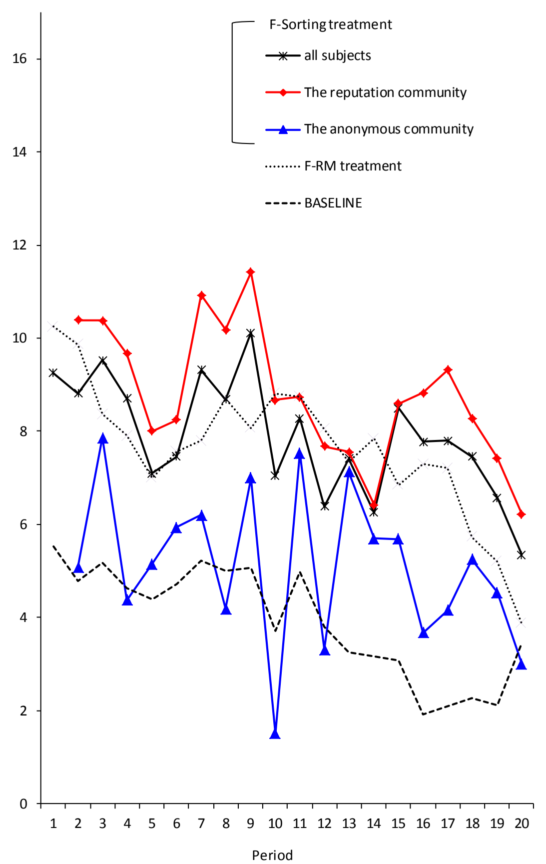



(a) Average contributions

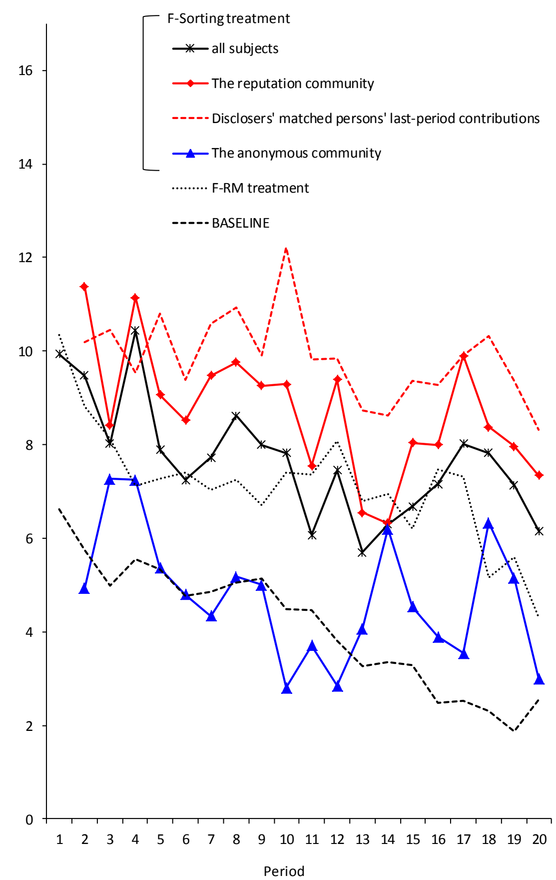
(b) Average beliefs on their matched partners' contributions

(c) The percentage of those who disclosed/ sorted into the reputation community

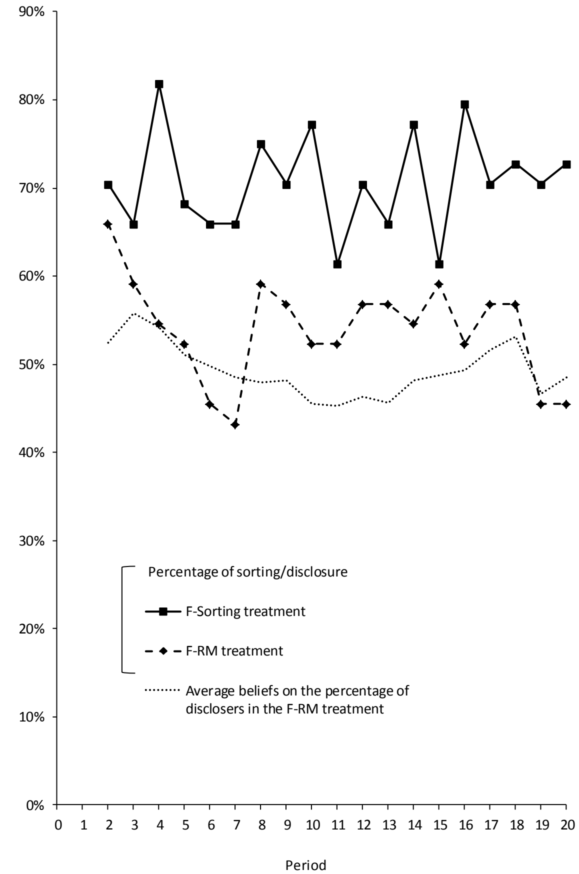
(I) C-Sorting and C-RM treatments




(a) Average contributions




(b) Average beliefs on their matched partners' contributions




(c) The percentage of those who disclosed/ sorted into the reputation community

(II) F-Sorting and F-RM treatments

*Notes*: The red dash line depicted in figure (b) shows the average of last-period contributions made by the disclosers' matched partners to the joint accounts in the C-Sorting (F-Sorting) treatment in Panel I (Panel II). See Appendix Figure C.1 for period-by-period diagrams of subjects' moving-in and moving-out of the reputation community in the C-Sorting and F-Sorting treatments, which shows that there are a stable number of events with moving-in and moving-out in every period in the two sorting treatments.

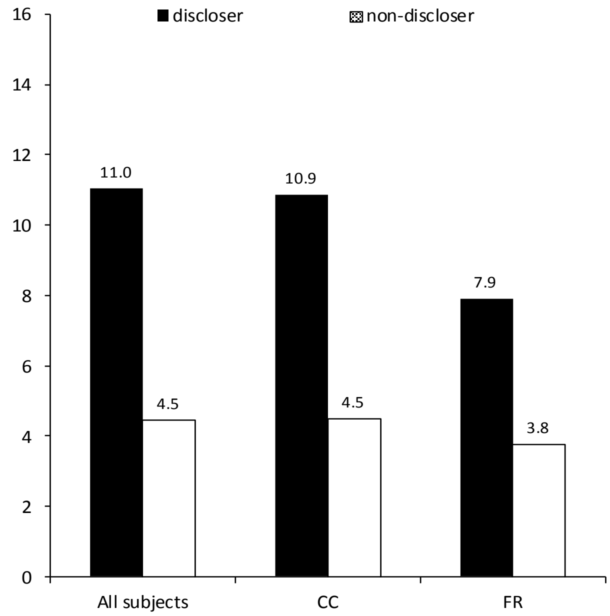
**Table 2.** *The Impact of Information Disclosure or Sorting on Cooperation*

| Independent Variable: | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| C-RM dummy {= 1 for the C-RM treatment; 0, otherwise}(#1) | 1.99<br>(1.39) | 1.91<br>(1.20) | 1.99<br>(1.39) | 1.91<br>(1.20) |
| C-Sorting dummy {= 1 for the C-Sorting treatment; 0, otherwise} (#2) | 5.19***<br>(1.38) | 5.57***<br>(1.52) | --- | --- |
| *C-Sorting reputation community dummy (#2a)* | --- | --- | *7.82****<br>*(1.39)* | *7.65****<br>*(1.47)* |
| *C-Sorting anonymous community dummy (#2b)* | --- | --- | *2.52**<br>*(1.30)* | *3.96***<br>*(1.79)* |
| F-RM dummy {= 1 for the F-RM treatment; 0, otherwise} (#3) | 3.57**<br>(1.52) | 3.74***<br>(1.24) | 3.57**<br>(1.52) | 3.74***<br>(1.12) |
| F-Sorting dummy {= 1 for the F-Sorting treatment; 0, otherwise} (#4) | 4.07***<br>(1.58) | 3.67***<br>(1.35) | --- | --- |
| *F-Sorting reputation community dummy (#4a)* | --- | --- | *5.27****<br>*(1.72)* | *5.43****<br>*(1.56)* |
| *F-Sorting anonymous community dummy (#4b)* | --- | --- | *.95*<br>*(1.57)* | *-.47*<br>*(1.78)* |
| Period Number {= 2, 3, …, 19, 20} | --- | -.16***<br>(.039) | --- | -.16***<br>(.039) |
| Constant | 3.91***<br>(1.24) | 5.70***<br>(1.06) | 3.91***<br>(1.24) | 5.70***<br>(1.06) |
| # of Observations | 380 | 380 | 529 | 529 |
| *p*-value (two-sided) for Wald tests to the following hypothesis: | | | | |
| $H_0$: (#1) = (#2) | .0003*** | .0028*** | --- | --- |
| $H_0$: (#1) = (#2a) | --- | --- | .0000*** | .0000*** |
| $H_0$: (#1) = (#2b) | --- | --- | .4763 | .1841 |
| $H_0$: (#1) = (#3) | .1459 | .0060*** | .1460 | .0061*** |
| $H_0$: (#1) = (#4) | .0735* | .0767* | --- | --- |
| $H_0$: (#2) = (#3) | .1274 | .1139 | --- | --- |
| $H_0$: (#2) = (#4) | .3284 | .1659 | --- | --- |
| $H_0$: (#3) = (#4) | .7001 | .9357 | --- | --- |
| $H_0$: (#3) = (#4a) | --- | --- | .2503 | .1599 |
| $H_0$: (#3) = (#4b) | --- | --- | .0448** | .0044*** |
| $H_0$: (#2a) = (#2b) | --- | --- | .0000*** | .0130** |
| $H_0$: (#4a) = (#4b) | --- | --- | .0001*** | .0000*** |
| $H_0$: (#2a) = (#4a) | --- | --- | .0575* | .1473 |

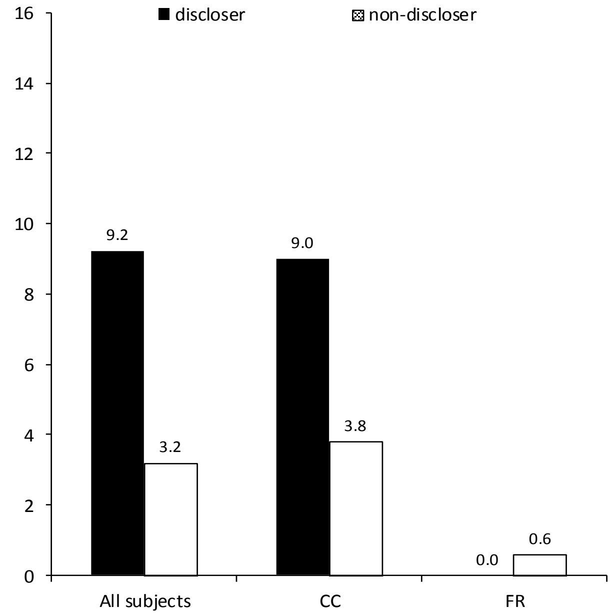
*Notes*: Linear regressions. Standard errors, in parentheses, were clustered by session. Random effects were included to control for panel structure because treatment dummies are included as regressors. Dependent variable is session-average contributions in period *t*. Observations in period 2 to 20 are used as data. The reference group is data from the BASELINE treatment. In addition to the Period Number variable, the interaction term between each treatment dummy and the Period Number variable was added as control variables in columns (2) and (4); because all the interaction terms, except the interaction term between variable (#4b) and Period number (for which the coefficient estimate is .13 and significant at the 5% level), failed to obtain significant coefficients we did not include them in this table to conserve space.

*, **, and *** indicate significance at the .10 level, at the .05 level and at the .01 level, respectively.

**Figure 3.** *Average Contribution Amounts by Disclosure Decision in the C-RM and F-RM treatments*

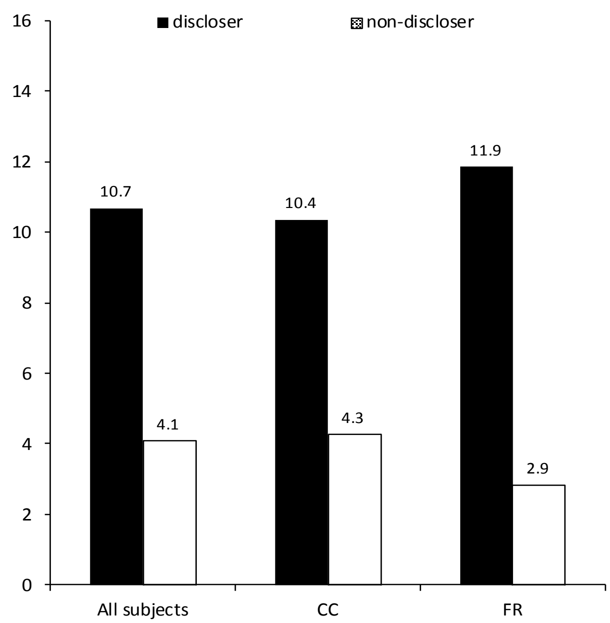



(a) Average across all periods before period 20

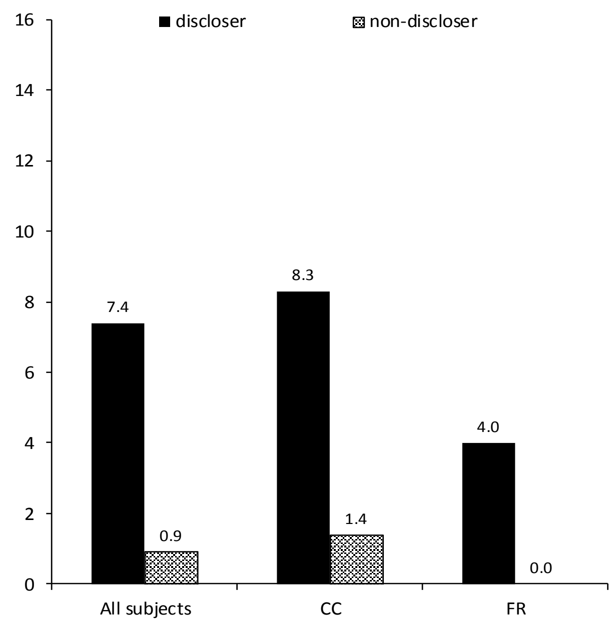



(b) Period 20 (final period of the experiment)

(1) The C-RM treatment




(a) Average across all periods before period 20




(b) Period 20 (final period of the experiment)

(2) The F-RM treatment

*Notes*: CC and FR refer to conditional cooperators and free riders, respectively. The 'all subjects' category includes also cooperation types other than conditional cooperators and free riders.

**Table 3.** *Subjects' Contributions, Beliefs, and Partners' Last-Period Contribution Amounts*

(a) Relationship between the Subjects' Contributions and Beliefs on their Partners' Contribution Amounts

Dependent variable: Subject $i$'s contribution amount to his or her joint account in period $t$, where $t \in \{2, 3, \ldots, 19\}$.

| Data: | The random-matching community | | The reputation community[#2] | | The anonymous community[#3] | |
|---|---|---|---|---|---|---|
| | (1)[#1] | (2)[#1] | (3) | (4) | (5) | (6) |
| Independent Variable: | C-RM | F-RM | C-Sorting | F-Sorting | C-Sorting | F-Sorting |
| (i) Subject $i$'s belief on his or her matched partner's contribution amount in period $t$ | .58*** | .49*** | .45*** | .55*** | 39*** | .26 |
| | (.027) | (.055) | (.046) | (.078) | (.045) | (.18) |
| Constant | 2.52*** | 4.22*** | 6.77*** | 4.03*** | 3.58*** | 4.36** |
| | (.16) | (.39) | (.56) | (.69) | (.32) | (.84) |
| # of Observations | 864 | 792 | 423 | 559 | 369 | 233 |
| F | 486.71 | 78.08 | 93.57 | 50.69 | 77.60 | 2.25 |
| Prob > F | .0002 | .0031 | .0023 | .0057 | .0031 | .2302 |

*Notes*: Linear regressions. Standard errors in parenthesis were clustered by session. Individual fixed effects were included to control for panel structure. Observations in period 20 were not included because the usual end-game defection was observed (Andreoni, 1988). The results are similar even if observations in period 20 are included (the results are omitted in the table to conserve space). See Appendix Table C.5 for results when individual random-effects ordered probit regressions with standard errors clustered by session were used to estimate the relationships.

[#1] We also ran the same regressions while also having a dummy variable which equals 1 (0) if subject $i$'s period $t$ partner did not disclose (disclosed) his or her period $t-1$ contribution amount as an independent variable. The analysis shows that the dummy variable fails to obtain a significant coefficient while independent variable (i) obtains a significantly positive coefficient (the size of the coefficient estimates are almost similar to the ones in the above table). The results are omitted to conserve space.

[#2] A small number of the disclosers were matched with non-disclosers as explained in the text. Results in columns (3) and (4) are similar even if we exclude the observations in which disclosers were matched with non-disclosers. The results are omitted to conserve space.

[#3] A small number of the non-disclosers were matched with disclosers as explained in the text. Results in columns (5) and (6) are similar even if we exclude the observations in which non-disclosers were matched with disclosers. The results are omitted to conserve space.

*, **, and *** indicate significance at the .10 level, at the .05 level and at the .01 level, respectively.

(b) Relationship between the Subjects' Belief Formation and Partners' Last-period Contributions

(b1) The C-RM and F-RM treatments

Dependent variable: Subject $i$'s belief on his or her matched person $j$'s contribution amount in period $t$, where $t \in \{2, 3, \ldots, 19\}$.

| Independent Variable: | C-RM (1) | F-RM (2) |
|---|---|---|
| (1 – the No Information dummy) × $j$'s period $t-1$ contribution amount | .32*** (.016) | .39*** (.075) |
| Constant | 4.84*** (.055) | 4.44*** (.52) |
| # of Observations | 864 | 792 |
| F | 388.97 | 26.45 |
| Prob > F | .0003 | .0142 |

*Notes*: Linear regressions. Standard errors in parenthesis were clustered by session. Individual fixed effects were included. The reference group is those whose matched partners did not disclose their last-period contribution amounts. Observations in period 20 were not included because of the strong end-game defection observed in the experiment. However, results are similar even if observations in period 20 was included (the results are omitted to conserve space). The No Information dummy equals 1 if subject $i$'s period $t$ matched person $j$ did not disclose his or her last-period contribution amount in period $t$; and 0 otherwise. See Appendix Table C.5 for results when individual random-effects ordered probit regressions with standard errors clustered by session were used to estimate the relationships. *, **, and *** indicate significance at the .10 level, at the .05 level and at the .01 level, respectively.

(b2) The Reputation Community in the C-Sorting and F-Sorting treatments

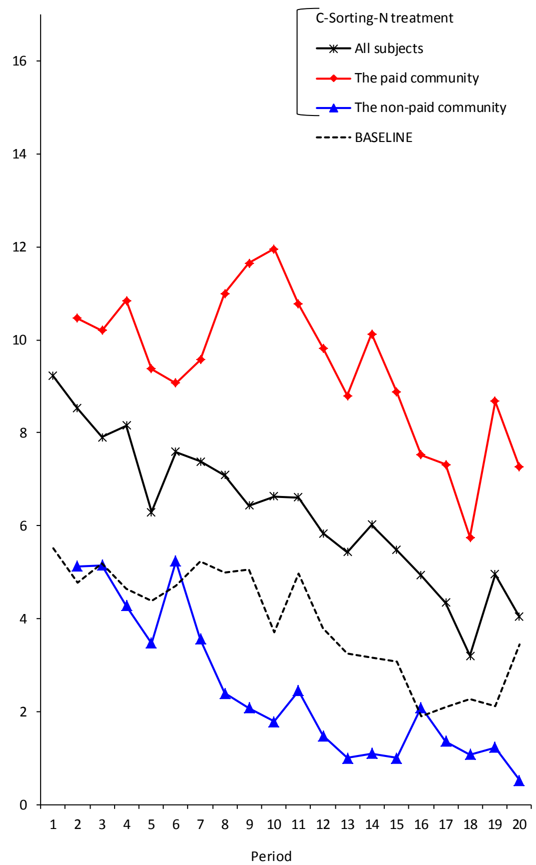
Dependent variable: Subject $i$'s belief on his or her matched discloser $j$'s contribution amount in period $t$, where $t \in \{2, 3, \ldots, 19\}$, in the reputation community.

| Independent Variable: | C-Sorting (1) | F-Sorting (2) |
|---|---|---|
| Subject $i$'s period $t$ matched discloser $j$'s period $t-1$ contribution amount | .48*** (.053) | .42*** (.031) |
| Constant | 6.03*** (.74) | 4.51*** (.33) |
| # of Observations | 384 | 520 |
| F | 82.56 | 181.71 |
| Prob > F | .0028 | .0009 |

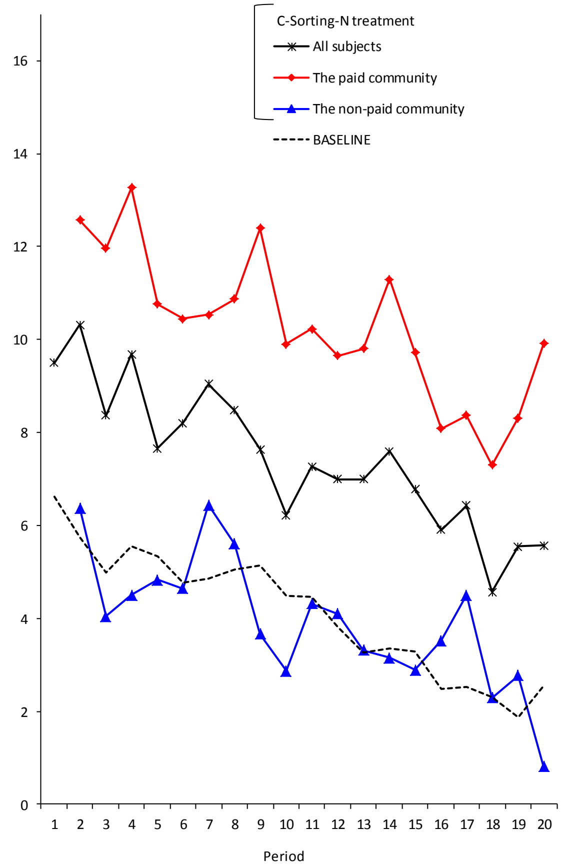
*Notes*: Linear regression. Standard errors in parenthesis were clustered by session. Individual fixed effects were also included. Only observations in which discloser $i$ was matched with another discloser $j$ were used in the regression analyses. Observations in period 20 were not included because of the strong end-game defection observed in the experiment. However, results are similar even if observations in period 20 was included (the results are omitted to conserve space). See Appendix Table C.5 for results when individual random-effects ordered probit regressions with standard errors clustered by session were used to estimate the relationships.

*, **, and *** indicate significance at the .10 level, at the .05 level and at the .01 level, respectively.

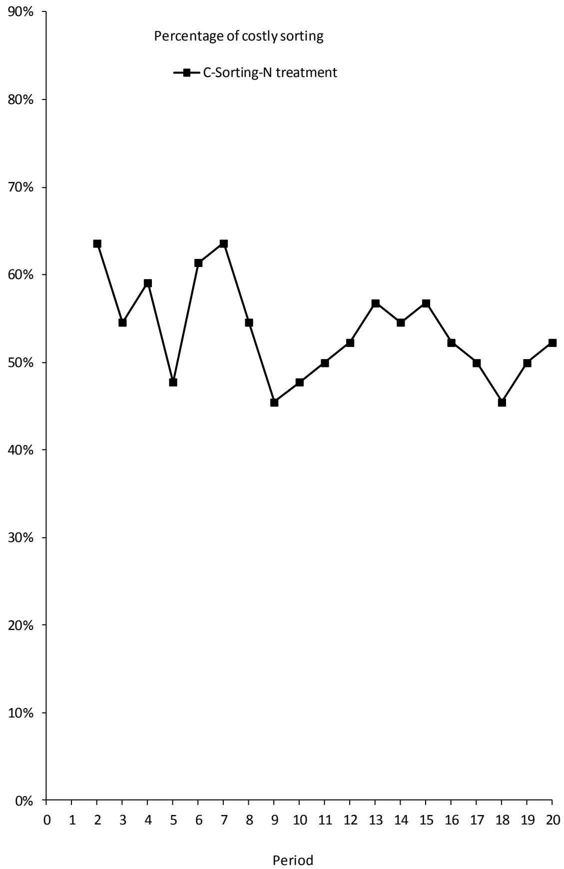
**Figure 4.** *Period-by-Period Average Contributions and the Percentage of the Subjects Who Paid a Fee to Join the Paid Community in the C-Sorting-N treatment*




(a) Average contributions

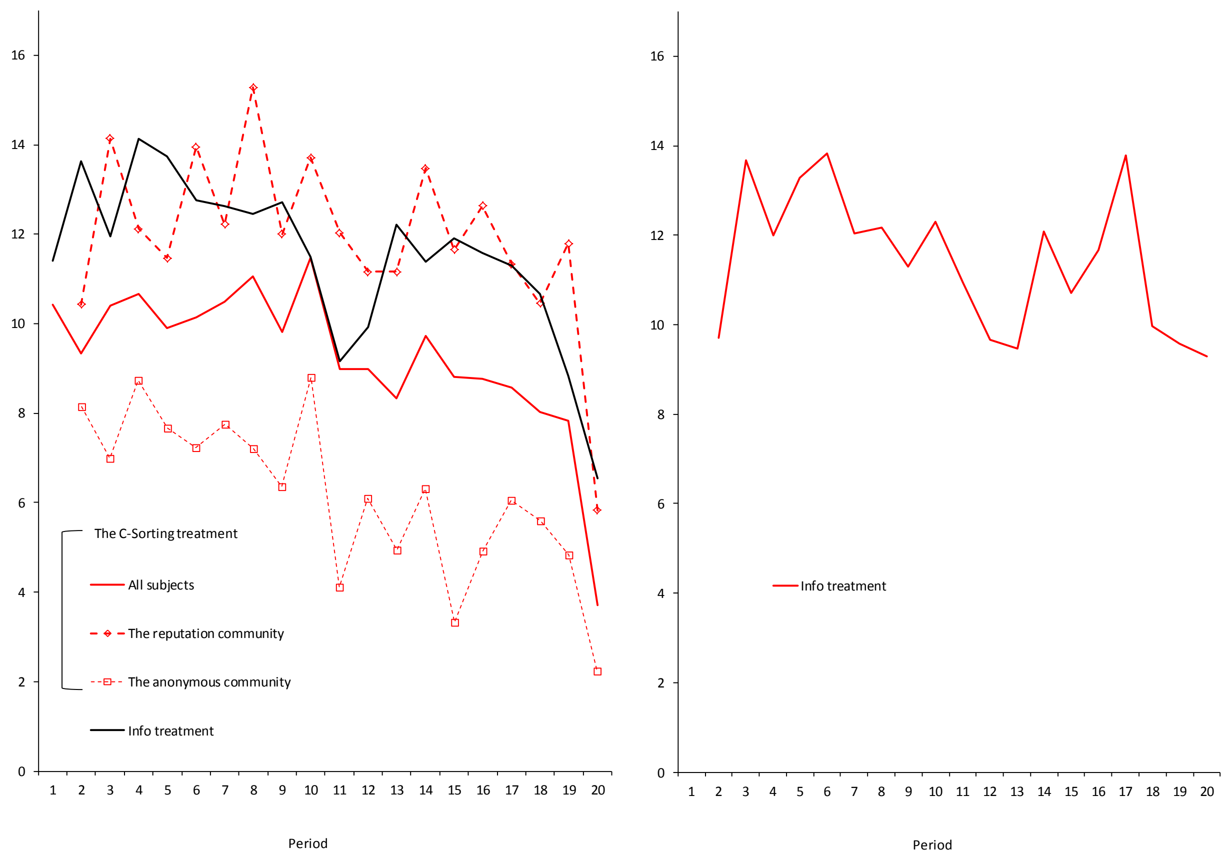



(b) Average beliefs on their matched partners' contributions




(c) The percentage of those who sorted into the paid community

**Figure 5.** *The Trends of Average Contribution in the Info treatment*




(a) Average contributions

(b) Average beliefs

*Notes*: 1. The average payoffs in the C-Sorting anonymous community and the Info treatment are monotonic transformation of average contributions, based on Eq. (1). The average payoffs in the C-Sorting reputation community are monotonic transformation of average contributions, based on Eq. (1), minus 1 (sorting cost). Thus, the average payoffs are lower in the C-Sorting reputation community than in the Info treatment by around 1 ECU.

2. As in other treatments, the average beliefs in the Info treatment transited similar to the average contributions.