Finding cooperators:
Sorting through repeated interaction

Mark Bernard Jack Fanning Sevgi Yuksel

April, 2017

Abstract

We present evidence from an indefinitely repeated gift-exchange game where market structures affect proposers' ability to punish uncooperative partners and their ability to sort between cooperative and uncooperative partners. Treatments vary by whether subjects can replace their partners, and if not, whether they can reduce their gift from one round to the next. Comparing treatments without contract restrictions, our replacement treatment is no different initially but has higher cooperation in the long run. Comparing treatments without replacement, our treatment with contract restrictions has lower cooperation initially but is no different in the long run. Neither of these findings are predicted by theories of repeated games based on the ability to punish, however, both findings are consistent with a simple sorting model.

This research was supported by: National Science Foundation Grant No. 1058380, and the Center for Experimental Social Science (CESS) at New York University. An earlier version was circulated under the title *The benefits of firing (and replacement): an experimental study.* We would like to thank Andrew Schotter, Magnus Johannesson, Ernst Fehr, David Cesarini, Mark Dean, Pedro Dal Bó, experimental group at NYU, participants in seminars and conferences at NYU, the Stockholm School of Economics, Brown, Oxford, FSU, BI Oslo, the University of Vienna, Maastricht University, CREED, EconCon 2011 and ESA 2011 for helpful discussions. We are especially grateful to Guillaume Fréchette for his patience and advice.

Mark Bernard: *Economic Analysis Unit, Swedish Post and Telecom Authority (PTS), Box 5398, 102 49 Stockholm. Email: mark.bernard@pts.se*,
Jack Fanning: *Corresponding author. Department of Economics, Brown University, Robinson Hall 303C, 64 Waterman Street, Providence, RI 02912. E-mail: jack_fanning@brown.edu*,
Sevgi Yuksel: *Department of Economics, University of California Santa Barbara, 13044 North Hall Santa Barbara, CA 93106. E-mail: sevgi.yuksel@ucsb.edu*

1. Introduction

Many studies in the experimental literature have found that repeated interactions generate higher levels of cooperation. A standard theoretical explanation for this result centers on the incentives created by the threat of punishment. An agent chooses to cooperate because she anticipates that a partner will punish her if she does not (e.g. by reducing her own cooperation in the future). Hence, according to the theory, differences in the ability to exercise repeated-game punishments entail differences in the potential for cooperation: cooperative opportunities increase with agents' ability to punish.

We conducted an experiment using a repeated gift-exchange game to study the importance of dynamic punishment strategies in establishing cooperation. The gift-exchange game is often used to model worker-firm relationships. A *proposer* (firm) makes a *transfer* gift (wage) to a *responder* (worker) and requests a *return* gift (effort); the responder then chooses the size of her return. The proposer's benefit from an increased return is larger than its cost to the responder, making mutual gift-exchange efficient. Given the game's asymmetry, we identify cooperation levels with the size of return gifts.

At first glance, the existing experimental literature on repeated gift-exchange games would appear to support the theoretical prediction that repeated-game punishments increase cooperation. Brown, Falk & Fehr (2004) (BFF) show that when a proposer can make gifts to one of many possible responders, the ability to identify partners from previous rounds increases cooperation. The authors explain this finding by highlighting that identification is necessary for effective punishment. With identification, proposers can punish the responder for non-cooperative behavior by terminating a relationship that otherwise offers future rents. Without identification, relationships (are very likely to) terminate regardless of behavior. Falk, Huffman & MacLeod (2008) (FHM) show that dismissal barriers, which prevent a proposer from terminating a relationship or reducing her transfer, lower cooperation.[1] Again, the authors explain this result by highlighting how dismissal barriers restrict the ability to punish.

On closer inspection, however, it is not clear whether the above results can be attributed entirely to differences in punishment threats. Both BFF and FHM simultaneously vary a proposer's ability to punish an uncooperative responder *and* her ability to find and develop a relationship with a cooperative partner. Without identification in BFF, it is impossible for proposers to continue relationships only with cooperative responders (regardless of why responders cooperate). This may lower aggregate cooperation even under the extreme assumption that all responders are behavioral types, unresponsive to punishment threats, who either cooperate (provide the requested return) or don't. Similarly, dismissal barriers in FHM prevent a proposer from replacing an uncooperative

[1] FHM's dismissal barriers apply only if a proposer interacts with a responder for two consecutive rounds, mirroring employment protection legislation in a labor market context.

responder type with a possibly more cooperative one.

We refer to a proposer's ability to distinguish between cooperative and uncooperative responders and make transfers only to the former, as sorting. Our experiment attempts to distinguish between a proposer's ability to sort and ability to punish, and see the extent to which each factor predicts cooperation. Our treatments independently varied proposers' ability to punish uncooperative responders, and their ability to replace them (the simplest form of sorting). A Baseline (B) treatment allows replacement options and flexible contracts; this resembles the baseline treatments of BFF and FHM. A Fixed Partner (FP) treatment retains flexible contracts but restricts proposers to making offers to a single responder. Finally, a Contract Restrictions (CR) treatment imposes a fixed partner and additionally proposers cannot reduce their transfers;[2] this resembles the dismissal barriers of FHM. In all treatments, after each round of gift giving, play continued for another round with a constant probability. We refer to each repeated interaction (of indeterminately many rounds) as a supergame.

The theory of repeated games which focusses on the ability to punish (we shall henceforth refer to this as *repeated game punishment theory*) does not predict any difference in cooperation between the Baseline and Fixed Partner treatments. The same maximal cooperation level can be obtained in both settings from the start of a supergame by threatening responders with a zero continuation payoff punishment if they do not cooperate. Indeed the set of possible equilibrium payoffs for a proposer–responder pair is *identical* in the two treatments. On the other hand, the theory does predict a clear difference in cooperation between these flexible contract treatments and the Contract Restrictions treatment. Contract Restrictions constrain punishments, destroying the incentive to cooperate; the unique sequential equilibrium involves no gift-exchange.

Alternatively, if the ability to sort between cooperative and uncooperative responders is instrumental in generating cooperation, we would expect different results between the Baseline and Fixed Partner treatments. In the Baseline treatment, uncooperative partners can be recursively replaced with possibly cooperative ones, allowing for a larger population of cooperative matches to emerge over time. By contrast, in both the Fixed Partner and Contract Restrictions treatments a proposer can only hope to interact with a single responder, limiting the potential for sorting. For example, under the extreme assumption that all responders are cooperative or uncooperative behavioral types, we might expect similar cooperation levels in these two treatments.

Our results can be summarized as follows: in the Baseline and Fixed Partner treatments, the average return is not significantly different in the early rounds of a supergame, but it is higher in the Baseline treatment in later rounds. By contrast, the average return in the Fixed Partner treatment is larger than in the Contract Restrictions treatment in

[2]In all treatments, transfers and returns are chosen from a range, as shown in Section 2.

early rounds of a supergame but is not significantly different in later rounds.

The comparison between the Baseline and Fixed Partner treatments, both in terms of initial and eventual cooperation levels, points directly to the importance of sorting. The finding that cooperation rates do not differ under the Fixed Partner and Contract Restrictions treatment in later rounds is more remarkable; it suggests that the threat of repeated game punishments may have played little role in determining long-run cooperation levels.[3]

To fully understand the roles of sorting and punishment threats, however, requires examining the dynamics of the differences in cooperation between the Fixed Partner and Contract Restrictions treatments. Why is cooperation under Contract Restrictions initially lower, but ultimately similar? Mechanically, this occurs because cooperation in the Fixed Partner treatment remains approximately constant, while it increases over the supergame under Contract Restrictions. Such increased cooperation under Contract Restrictions is explained by the predominance of "starting small" strategies. Proposers initially made small transfers and requested modest return gifts. When this return was provided, proposers subsequently increased the size of both their own transfer and their requested return.

Starting small can be explained as a form of sorting. It allows a proposer to distinguish between cooperative and uncooperative responders at low stakes, and increase the stakes only with the former. Starting small would seem particularly important for sorting under Contract Restrictions when the inability to reduce transfers increases the cost of facing an uncooperative partner considerably. In the theory section we present a simple model with behavioral responder types, developed *ex-post*, to illustrate how sorting can provide incentives for proposers to start small in all treatments. The model further shows that those incentives are greatest under Contract Restrictions.

It might seem that starting small can also provide incentives for responders. A proposer can "punish" a responder who does not cooperate by failing to increase her transfer. Countering that interpretation, however, note that under repeated game punishment theory starting small cannot incentivize cooperation under Contract Restrictions, given an upper bound on gifts.[4] More generally, even if starting small could provide short run incentives for responder cooperation, it is difficult to imagine that these are of comparable strength to the incentives available through punishment in our Fixed Partner treatment. Ultimately, we believe that sorting offers the most coherent explanation of both the Fixed Partner-Contract Restrictions comparison and the Fixed Partner-Baseline comparison.[5]

[3]Experimental public-goods games have traditionally implemented punishment by allowing a subject to choose to sacrifice money to lower the payoff of other parties. Our repeated game offers less explicit punishments, in that proposers can withhold future transfers that were otherwise expected (this distinction can also be seen in the Charness & Rabin (2002) model).

[4]A responder won't make a return gift once a proposer's transfer reaches the maximum, so a proposer will never increase gifts that far. This logic unravels so that a proposer would never offer positive transfer.

[5]It certainly seems difficult to argue that there was little difference between the Fixed Partner and

To the extent that sorting drives our results, they seem to offer an important counterpoint to the dominant theoretical perspective that repeated interactions foster cooperation primarily (or only) because of the punishment opportunities they create. While there is a theoretical literature on sorting within repeated games, considering in particular the incentives to start small (e.g. Watson (1999)), this is still far from the mainstream. Our results suggest that finding the "right" partner could be as important, if not more important, than providing sufficient incentives to an inherently self-interested partner in establishing a cooperative relationship. This conclusion seems to resonate beyond our experiment. Indeed, it broadly conforms to our own life experiences: we look to form friendships, co-authorships and business ventures with people who would help us even if it were not in their narrowly defined self-interest, and are nervous of trusting strangers who have not yet earned that trust, having been burned before.

Our results may have importance for institutional design. For instance, they suggest extensive screening to get the "right sort" of person into a job might be more important than the incentives that person has once there. Institutions like employment protection would seem likely to be inefficient, less because they discourage hard work, but because they prevent firms from substituting hard working people into jobs (sorting through replacement), discourage firms from giving workers big responsibilities and pay at the start of their careers (sorting through starting small) and discourage hiring in the first place because of the extra costs associated with "bad apples". Temporary contracts and internships, by contrast, may play a vital role in enabling effective sorting.

Our paper is most closely related the repeated gift-exchange literature and specifically to BFF and FHM.[6] Those papers demonstrate that the inability to identify previous partners and dismissal barriers lower cooperation.[7] We help shed light on why that occurs, in particular highlighting sorting. Both BFF and FHM find some evidence of proposers starting small, which is central to our explanation of the data. However, they do not directly propose sorting as an explanation.

In addition to addressing a different substantive question, we make two methodological innovations with respect to this literature. First, random termination of the supergame means our proposer–responder relationships have a stationary *indefinite* horizon instead of a non-stationary finite one. This allows for cooperation as an equilibrium outcome

Contract Restrictions in terms of the ability to punish, but at the same time the Baseline treatment offered larger punish threats (later in the supergame).

[6]This literature built on a series of experiments on one-shot gift-exchange games (e.g. Fehr et al. (1993, 1998)) which found significant gift-giving in contrast to subgame perfect predictions (assuming no other regarding preferences).

[7]BFF also shows that the gains of trade are shared equally within the gift-exchange setting, but not when contracts are third-party enforceable. FHM also shows that if proposers can pay bonuses (after a responder's return gift), dismissal barriers impose little efficiency cost. Their dismissal barriers also operate only if a proposer makes a transfer to the same responder in two consecutive rounds. They show that this causes proposers to avoid such long-term relationships.

even when all agents are selfish profit maximizers, bringing us closer to the theoretical relational contract literature (e.g. see MacLeod & Malcolmson (1989))).[8]

A second innovation for this literature is to have subjects play many supergames as opposed to one. This allows subjects to learn about the supergame environment. We then analyze data from only the last four supergames (subjects played at least seven). Engle-Warnick & Slonim (2004), Dal Bó (2005) and Dal Bó & Fréchette (2011) highlight the importance of learning about supergames. For instance, Dal Bó (2005) shows that cooperation rates in a prisoners' dilemma decreased significantly over a series of finitely repeated supergames but grew in indefinitely repeated games of the same expected length, in line with repeated game punishment theory. By allowing for learning we hope to observe behavior closer to a steady state (or equilibrium) that might exist outside the laboratory.

Finally, note that our results highlighting the importance of sorting is not in direct opposition to the consistent finding in the broader experimental literature on repeated games, that repeated interaction increases cooperation rates even when players cannot replace their partner and starting small is impossible (e.g. in the prisoners' dilemma). Even in the context of our simple sorting model, under the extreme assumption of behavioral responders types, repeated interaction should increase cooperation (compared to one-shot) when proposers can make only a fixed positive transfer or zero transfer and cannot replace responders. This is because repetition increases a proposer's incentive to risk making an initial gift to initiate a cooperative relationship, as there is a larger upside payoff from finding a cooperative responder. Of course, this is not to claim that those past results are in fact explained by sorting, but rather to highlight that sorting may be a general channel through which repeated interactions can foster cooperation.

1.1 Additional related literature

The importance of partner selection in generating cooperation has been emphasized in a few different contexts. Some of these papers focus on how exclusion (from a potentially cooperative interaction) can be used as a disciplinary measure against defectors (See Coricelli et al. (2004), Charness & Yang (2014), Croson et al. (2015))

A closer look at how cooperation arises in such setting also reveals a great deal of heterogeneity in willingness to cooperate. Several experiments highlight the ability to sort in terms of inherent cooperativeness via voluntary association or endogenous group formation to be critical in generating cooperation. Roe & Wu (2009) classify subjects into selfish and cooperative types and provide evidence in the context of a finitely repeated labor market of selfish types mimicking cooperative types in early rounds when individual reputations can be tracked.[9] Page et al. (2005) show in a public goods setting that when

[8]Note hat BFF and FHM account for cooperation in their finite horizon game by necessarily assuming the existence of cooperative (perhaps fair) responder types which selfish responders imitate.

[9]This is consistent with strong end-game effects in finitely repeated prisoner's dilemma experiments.

subjects can influence whom they interact with, they sort into groups displaying different degrees of cooperativeness. Gächter & Thöni (2005) classify subjects into types based on cooperative attitudes and examine how sorting subjects based on their type affect behavior in a public goods game. They argue that cooperation *only* arises among "like-minded" cooperators. Castillo & Petrie (2010) and Charness et al. (2014) have experimental designs with group formation in the context of a public goods game. Both papers explore the effects of endogenous sorting on cooperation and the role social identity plays in the formation of groups. Examining purely preference-based sorting, Lazear et al. (2012) show that, in the context of a dictator game, allowing subjects to avoid environments in which sharing is possible significantly reduces aggregate sharing, revealing the existence of a "reluctant" sharing type. Aggregate sharing increases again when it is subsidized, but less is shared on average by those choosing to share. Another experimental study on the emergence of long-term relationships that actually builds on the very same baseline design as ours is Eriksson & Villeval (2012). The authors describe how proposers use symbolic rewards as a coordination device to initiate relational contracts. Kurzban et al. (2001) and Kurzban et al. (2008) investigate the role of incremental commitment for cooperation, the first in a real-time public goods game and the second in an investment game. Both papers highlight how establishing trust in small increments can be effective in establishing cooperation.

In the context of a repeated trust game, Rigdon et al. (2007) rank subjects according to a "trust score" from most trusting to least, and match trusting proposers with trustworthy responders. Over the course of a session cooperation increases among players with high "trust scores" while it decreases for others, suggesting that sorting can help sustain trust and reciprocity.

Similar patterns are observed with the prisoners's dilemma. In Hauk & Nagel (2001), subjects are offered a choice between a sure outside option and playing a finitely repeated prisoner's dilemma. This creates sorting, as subjects intending the defect in the game are more likely to take up the outside option. Among players who choose to play the game cooperation is higher relative to the baseline with no choice. In Grimm & Mengel (2009), players choose between two groups where different groups play prisoner's dilemmas with different values attached to defecting. Choosing across these games is used to signal willingness to cooperate, and creates endogenous sorting between cooperators and defectors. Janssen (2008) show that cooperation can arise, even in non-repeated interactions between unrelated agents, when the agents have the ability to detect the trustworthiness of other agents. Similarly, in Aimone et al. (2013), subjects voluntarily join groups that provide *reduced* rates of return on private investment. This is interpreted as a "sacrifice" mechanism used to facilitate endogenous sorting. In contrast to our experimental design,

(High levels of cooperation followed by sharp decline in the last few rounds of the supergame.) See Embrey et al. (2016) for an overview of this literature.

all these papers investigate sorting in settings where it is very difficult (or theoretically impossible) to provide incentives to support cooperation otherwise.

Andreoni & Samuelson (2006) conduct an experiment on starting small using a two round prisoners' dilemma in which the stakes (payoffs for a given action profile) in the two rounds differ. Keeping the same average stakes over the two rounds (so there is no efficiency cost to starting small) they find that cooperation is higher when stakes are backloaded. Andreoni et al. (2016) shows that subjects learn to choose this game structure after repetition of the supergame. The two sided nature of the prisoner's dilemma and the increased strategic uncertainty implied make a direct comparison to our experiment difficult. However, the results seem consistent with the importance of sorting and subjects who want to cooperate if and only if their partner does likewise.

The personnel economics literature has studied the importance of sorting in labor markets, as well its interaction with incentives. For instance Lazear (2000) shows that the productivity gains from performance related pay at a car windshield repair company, are divided equally between increased output due to incentives, and greater retention of more productive (over less productive) workers. A major difference between this and our experiment is that our subjects have the same inherent productive capacity for making gifts, but only some choose to use it to build mutually beneficial relationships. Even in a seemingly pure moral hazard setting, sorting matters.

2. Experimental Design

Our experiment implemented an indefinitely repeated gift-exchange game. We have three treatments: Baseline (B), Fixed Partner (FP), and Contract Restrictions (CR). Each experimental session consisted of 20 subjects, with the exception of one session with 17 subjects, and was conducted as follows:

A subject was randomly assigned to be a proposer or responder. There were 7 proposers and 13 responders in each treatment.[10] [11] Proposer/responder assignments were kept constant throughout the session. Each session consisted of a series of indefinitely repeated gift-exchange games (supergames). Within each supergame each subject was assigned an ID number, however, these were randomly reassigned between supergames. We refer to each interaction between a proposer and responder *within* a supergame as a *round.*

We experimentally induced an indefinite horizon by having the game continue to the next round with a fixed probability $\delta = 0.8$. This implied that the expected length of a supergame was 5 rounds. Such a δ allowed us to balance the need for a significant future

[10]In one FP session fewer subjects were available and this was reduced to 6 proposers and 11 responders.

[11]In two sessions, one B and one FP, we experienced some issues with computers crashing in the middle of supergames. Our procedure has been to drop all data for affected subjects from the point of any crash until the end of the supergame, when programs were reloaded.

with the need to run several repetitions of the supergame.

Each round of the supergame consisted of three stages. In stage one, proposers made offers to responders, stipulating a transfer (a gift) and a requested return. In stage two, responders chose which offers to accept. In stage three, those responders who had accepted offers made their return (gift) decisions.

An offer consisted of a transfer t, and a desired return $\tilde{r}$. The transfer was chosen in experimental points from $\{0, 1, ..., 40\}$, and the desired return level was from $\{0, 1, ..., 10\}$. The transfer was paid to a responder who accepted the offer regardless of whether that responder actually made the return requested by the proposer.

Offers could either be *public*, visible to all responders, or *private*, visible to only one responder. Proposers also had the ability to block certain responders from accepting public offers; this blocking would only be observed by the blocked responder. This allowed the proposer to ensure that it did not (re)match with a particular responder. An on-screen history box detailed previous offers and outcomes within the supergame, including proposer–responder ID numbers.

A proposer could only make a private offer to a responder she had been matched with in the previous round. This implies that in the first round of any supergame all offers were necessarily public. Additional restrictions on the types of offers proposers could make after round one differed by treatment variation. In treatment B a proposer could always choose whether to make a public offer or a private offer (if matched). In treatment FP and CR, however, a proposer could *only* make private offers if she was matched with a proposer in the previous round. In this sense partners were fixed in FP and CR. Furthermore, in CR, a proposer had to offer a proposer she was matched with at least as large a transfer as in the previous round.

Stage two started with all responders and proposers observing all public offers. After this, public offers became available for unmatched responders to accept. A responder who received a private offer also saw that offer in a separate box and could accept or reject it. If a responder accepted an offer (by clicking on it), that proposer and responder were *matched*, and if the proposer's offer was public, it immediately became unavailable to all other responders. Given that there was an excess supply of responders, matching resulted in a pool of at least 6 unmatched responders.

There was a small additional difference in treatments regarding what happened if a responder rejected a private offer. In treatments B and CR, the proposer and responder became *unmatched* (and so the responder could accept any available public offer). In treatment FP, however, the proposer and responder remained matched and waited until the next round. In theory this meant responders could choose to become unmatched in CR but not in FP (this design choice was made to make CR more closely resemble the institution of employment protection legislation). However, given that there were more responders than proposers, there was little outside option for proposers who rejected private offers: only 3% of offers after the first round were public in CR (of which 38% offered a zero transfer). It is no surprise therefore that less than 1% of private offers with a positive transfer were rejected in CR (similarly less than 1% of such offers were rejected in B). Importantly, because *proposers* cannot choose to become unmatched and transfers cannot be reduced, there was no means to punish responders in CR.

The third stage consisted of those responders who had accepted offers choosing how much to return. A responder chose how many points to return back to the proposer, $r \in \{0, 1, ..., 10\}$. Any return gift was quadrupled.

Proposers and responders both had an initial endowment of 10 points in each round. This gives us the following payoff functions: $\pi_P = 4r - t + 10$ and $\pi_R = t - r + 10$. Also this implies that total surplus generated in each round of a proposer–responder relationship equaled $\pi_P + \pi_R - 10 - 10 = 3r$, where we subtracted endowments because they were given whether or not a relationship was formed. Because of this we use r as a summary measure of the degree of cooperation between subjects; any $r < 10$ is inefficient. By allowing responders to return 0, relationships are more efficient than unmatched pairs only if the responder chooses to cooperate.

Table 1: Summary of treatments

	Flexible Partner	Fixed Partner
Flexible Transfer	*Baseline (B)*	*Fixed Partner (FP)*
Downward Fixed Transfer	-	*Contract Restrictions (CR)*

At the end of each round of each supergame, all subjects observed their own payoff and their partner's payoff if matched. Information about other matched pairs was not available. Table 1 summarizes our three treatments.[12,13]

Our sessions lasted for up to two hours (including instructions). In order to allow many supergames to be played, we had time limits on subjects' decisions at different stages of the game, such that the each round lasted up to 100 seconds.[14] Default decisions applied if no

[12] We also elicited subjects' beliefs (using a quadratic scoring rule, Murphy & Winkler (1970)) but found these did not help explain our results and so exclude them from our analysis for the sake of brevity. In every treatment, and every round of a supergame we asked proposers to guess what their average payoff per round would be in the *next* supergame. Because IDs were randomly rematched between supergames this could plausibly be viewed as a proposer's "outside option" in B. We also asked responders to guess their currently matched proposer's answer to the above question. Finally, we asked proposers what return they expected from their currently matched responder. Payoffs for belief elicitation were small relative to the stakes of the actual game (0-9 points per supergame compared to 0-250 points in a supergame of 5 rounds).

[13] We provided subjects with payoff tables detailing the profits to a proposer and responder pair as a function of the transfer and return (principally to help facilitate belief estimates).

[14] We allowed 20 seconds longer for the first 5 rounds of the experiment while subjects familiarized themselves with the game interface.

decision was taken.[15] The first supergame to end after 70 minutes of play determined the end of the session.[16] Thus, the total number of supergames and total rounds (cumulative over the supergames) in each session was randomly determined. The minimum number of supergames in a session was 7, the maximum 11 (median 9.5) while the minimum number of total rounds in a session was 40 and the maximum 58 (median 50).

We ran 4 sessions of each of our 3 treatments. The experiments took place at the Center for Experimental Social Science (CESS) at New York University. Subjects were drawn from undergraduates from all majors at NYU who had signed up to take part in economic experiments, through an online recruitment system. The experiment was programmed and conducted with the z-Tree software (Fischbacher (2007)).[17] Subjects were paid for their winnings from all supergames as well as from beliefs, at an exchange rate of 50 points per \$1. Participants' earnings averaged \$27.

3. Theory

3.1 Punishment and responder incentives

In an indefinitely repeated game, a proposer's threat to withdraw future gifts can incentivize a responder to make a positive return in every round of a sequential equilibrium, even when all subjects are selfish profit maximizers. This is in contrast to the prediction in finitely repeated games where backward induction requires zero gift-exchange in all rounds.

A responder's supergame payoff $V_R(r)$ from making a return of size r is composed of her current round payoff $t - r$ plus an expected continuation payoff conditional on her action.[18] When a proposer has the ability to terminate the relationship (either through replacement or lowering of transfers to zero from then on), she can use zero continuation payoff as a punishment threat in response to an uncooperative responder (who does not make the requested return, $\tilde{r}$). Let δv be the responder's discounted expected continuation payoff if she cooperates, then the incentive constraint for the responder not to shirk is:

[15] These defaults applied very rarely (<1%), with most decisions taken well before the time limit. Its primary use was in the CR treatment, where the previous round's offer was made if the proposer did not specify a new one. Also, failure to specify a return in time led to a return of 0 being recorded.

[16] The continuation probability for the last supergame remained the same at 80 percent. Thus, the exact length of a session depended on the draw of random numbers for the last supergame, which we did not know beforehand.

[17] Instructions and other materials are available at: https://sites.google.com/a/brown.edu/jfanning/$https://sites.google.com/a/brown.edu/jfanning$. In addition to written instructions being read out loud, a short Power Point presentation detailed key features of the computer interface, and was followed by a computerized test of subjects understanding of the game and payoffs. Unfortunately, one of the presentation slides contained a wording mistake: in one instance, we referred to responders as "workers". The actual computer interface did not contain this mistake, with all interaction neutrally worded. This mistake was first pointed out to us by an anonymous referee. No subject ever mentioned this glitch when allowed to ask questions, which makes us hope it escaped their notice.

[18] We do not need to consider payoffs from endowments, which cannot affect incentives.

$$V_R(\tilde{r}) = t - \tilde{r} + \delta v \geq t + 0 = V_R(0)$$

Assuming this is achieved through stationary strategies, $v = V(\tilde{r})$, the condition reduces to $t \geq \tilde{r}/\delta$. To support the maximum return of 10 with $\delta = 0.8$ requires $t \geq 12.5$. This analysis confirms that the simple punishment strategy of terminating a relationship can incentivize efficient cooperation in our B and FP treatments.

In fact, not only do both these treatments support the maximum possible return but for any given proposer–responder pair, they generate the same equilibrium payoff set. In particular, in both treatments there are equilibria where either proposers or responders obtain the entire gains from cooperation. Even when replacement is possible there is always an equilibrium where *any* matched responder will return 10 for a transfer of 40, but will otherwise return zero. The maximum return can also be sustained as a renegotiation proof equilibrium in both treatments.[19]

In contrast to this extensive equilibrium multiplicity, in the CR treatment, there is a unique equilibrium prediction of zero transfer and return. To see this, notice that because transfers are downward sticky and bounded between 0 and 40, the responder will certainly return 0 if transfers ever reach 40. But given this, the proposer would never raise transfers so high; and thus responders necessarily return 0 if transfers ever reach 39. Iterating this argument produces the result. [20]

Summary (Repeated game punishment theory): Simple punishment strategies such as termination of a relationship can sustain (equally) high levels of cooperation in both B and FP. In the absence of such a punishment strategy, the unique equilibrium prediction is no cooperation in CR.

3.2 Sorting

When designing our experiment we hypothesized that cooperation might be higher in B than in FP because proposers had greater ability to sort. Uncooperative responders could be recursively replaced with (possibly) cooperative ones leading to a more cooperative population over time. Our experimental results were consistent with this and also showed that many proposers adopted starting small strategies, particularly in the CR treatment. We then realized that these starting small strategies could be used to sort workers. Starting small allows a proposer to distinguish the cooperative tendencies of responders at low stakes, reducing the cost of facing an uncooperative partner. This would seem particularly important under CR where proposers cannot reduce transfers.

[19] Assume that renegotiation only takes place at the start of each round. Proposers can punish uncooperative responders with a zero transfer for one round, while continuing to demand a return of ten.

[20] In fact, since short-term (one-round) incentives require $\Delta t \geq r$, unraveling happens in at most four reasoning steps at $r = 10$, for any strictly positive initial transfer, and even in two steps when the initial transfer is $t = 25$, the one-shot equal surplus split.

Below we outline a very simple model, developed *after* running our experiment, to formalize these ideas. We emphasize that we do not claim our model can fully capture all subject behavior in our experiment; our objective is merely to highlight how differences in cooperative tendencies and the ability to sort may cause differences in cooperation levels in our treatments.

We introduce heterogeneity of cooperative tendencies in the simplest possible way. We assume that there exists a fraction p of "good" responder types who reciprocate gifts by returning a share s of any transfer made to them (which means that $r = st$) and a fraction $(1-p)$ of "bad" responder types who always make a return of zero. To streamline the analysis, we assume $s = \frac{2}{5}$, which ensures that any gains from trade are shared equally within a relationship.[21] Note that, in this model, all responders are assumed to be behavioral - they do not respond to incentives created by punishment strategies.

We first simplify a proposer's action space to the possible transfers $\{0, \bar{t}\}$ where $\bar{t} > 0$. In the FP treatment, a proposer's supergame payoff from a Grim Trigger strategy, which makes transfers of $\bar{t}$ initially but reverts to zero if the responder's type is revealed to be bad, is given by:

$$V_P^{FP}(GT) = p\frac{4s\bar{t}}{1-\delta} - (1-p)\bar{t}$$

Substituting in for s and δ this strategy is profitable for the proposer (compared to never making offers) if $p \geq \frac{1}{9}$. The same type of calculation can also be done for a one shot game where making a transfer of $\bar{t}$ is profitable for the proposer only when $p \geq \frac{5}{13}$.[22] This shows that even when responders are behavioral, and don't respond to punishment threats, repeated interaction can increase the likelihood of cooperation by increasing the upside from finding a cooperative partner.

We next allow transfers of size $\{0, \underline{t}, \bar{t}\}$, where $\bar{t} > \underline{t} > 0$. A proposer can now start small, offering a transfer of $\underline{t}$ to start with, and reducing this to zero conditional on facing a bad type but increasing it to $\bar{t}$ conditional on facing a good type. This gives the following supergame payoff:

$$V_P^{FP}(SS) = p4s\left(\underline{t} + \frac{\delta\bar{t}}{1-\delta}\right) - (1-p)\underline{t}$$

This is positive whenever $p \geq \frac{5\underline{t}}{13\underline{t}+32\bar{t}} \in (0, \frac{1}{9})$, meaning that attempting to build a cooperate relationship is always profitable for sufficiently small $\underline{t}$. Comparing $V_P^{FP}(SS)$ to $V_P^{FP}(GT)$, starting small is more profitable whenever $p \leq \frac{5}{13}$. The benefit to starting

[21] Formally, s solves $t(1-s) = 4st - t$. This assumption is not essential to our qualitative predictions, but we find substantial support for this particular sharing rule in the data.

[22] In a one shot game, a proposer's profits from making a transfer of $\bar{t}$ are $V_P^{OS}(\bar{t}) = p4s\bar{t} - (1-p)\bar{t}$.

small is reduced exposure to bad types, the cost is reduced cooperation with good types.

The analysis for treatment B is very similar. The grim trigger and starting small strategies now additionally involve replacing any bad types with unknown types from a pool of unmatched responders (which is assumed to be infinitely large). We find that the cutoff values of p which determine the profitability of grim trigger and starting small are exactly the same as in FP.[23] The key difference between the B and FP treatments is instead the fraction of relationships which are functioning. The fraction of matched responders who cooperate in FP is p in all rounds, but it is $1-(1-p)^n$ in round n of the supergame in treatment B. Replacement progressively ensures that more good types are matched over time.

We next consider the CR treatment. The supergame payoff for a strategy which starts with a transfer of $\bar{t}$ is simply a scaled up version of one shot payoffs $V_P^{CR}(\bar{t}) = \frac{p4s\bar{t}-(1-p)\bar{t}}{1-\delta}$ (given that transfers cannot be cut to bad types). This is positive only if $p \geq \frac{5}{13}$. A proposer's supergame payoff from starting small is now:

$$V_P^{CR}(SS) = p4s\left(\underline{t} + \frac{\delta\bar{t}}{1-\delta}\right) - (1-p)\frac{\underline{t}}{1-\delta}$$

This is positive if $p \geq \frac{25\underline{t}}{33\underline{t}+32\bar{t}} \in (0, \frac{5}{13})$. Comparing $V_P^{CR}(SS)$ and $V_P^{CR}(GT)$ starting small is more profitable whenever $p \leq \frac{25}{33}$. This shows that there are greater incentives to start small under CR than under FP and B (where starting small is preferred only if $p \leq \frac{5}{13}$) because the costs of facing a bad type are much higher in CR and starting small minimizes them.

Summary (Sorting): Cooperation levels should be the same in B and FP in early rounds of a relationship, but higher in B in later rounds (because bad types are replaced with good types). Greater incentives to start small imply lower cooperation in CR than FP in early rounds of a relationship (for $p \in (\frac{5}{13}, \frac{25}{33})$), but the same level of cooperation in later rounds.

4. Results

Our results are divided into several subsections. First, we show how cooperation levels are affected by our treatment variations and argue that these are consistent with the importance of sorting. We then present further evidence supporting a sorting explanation, in particular, the incidence of replacement and starting small strategies.

Before beginning our analysis, we emphasize that we only consider data from the last

[23]Supergame payoffs in B are then equal to those in FP plus an extra continuation payoff of size $V_P^B(GT)$ or $V_P^B(SS)$, which is realized probability $\delta(1-p)$. This simply scales payoffs upwards so that $V_P^B(GT) = \frac{V_P^{FP}(GT)}{1-\delta(1-p)}$ and $V_P^B(SS) = \frac{V_P^{FP}(SS)}{1-\delta(1-p)}$.

four supergames of each session, to allow subjects to first learn about the supergame environment. Subjects played between 7 and 11 supergames in each session, although the last four supergames represent 49% of the observations in our database. We discuss how learning affects our results in section 4.5, which is the only time in our analysis where we use data prior to the last four supergames.

4.1 Average cooperation

Figure 1 illustrates the evolution of cooperation *within* a supergame, plotting the average return by round for the first 6 rounds in our three treatments. It shows that there was higher average levels of cooperation in B than in CR in all rounds, but comparing B and FP and then FP and CR is a more nuanced operation. There was little difference between B and FP initially, but after several rounds, cooperation in B was higher. By contrast, cooperation started at a higher level in FP compared to CR, but it rose in CR and fell slightly in FP, meaning there was little difference between the treatments by round 6. The figure also includes 95% confidence intervals for these averages derived from 6 round specific regressions of return on three treatment dummies (for the last four supergames), clustering at the level of a session.

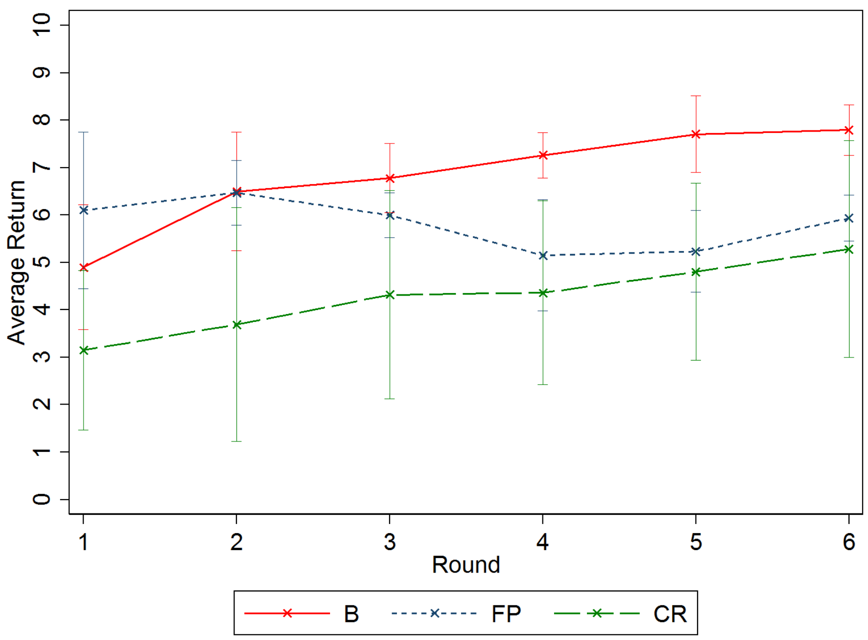

Figure 1: Average return over the course of a supergame.

Table 2 reports round specific OLS regressions of return on treatment dummies when comparing a single pair of treatments at a time. In the first row, first column we regress return in round 1 on an FP dummy and a constant. We report the coefficient on the FP dummy as well as standard errors (in parentheses) clustered at the session level (throughout the paper we *always* report standard errors clustered on the session level). Columns

2 through 7 report similar regressions for round 2, 3, 4, 5, 6 and for all rounds. Row two reports 7 further regressions of return but now using data from B and CR and a B dummy. Row three reports 7 more regressions of return using data from FP and CR and an FP dummy. These regressions involve up to 4 data points associated with each proposer (from 4 supergames).

The first row of the table shows that although the difference between B and FP is not statistically significant at the beginning of a supergame, it is at the 1% level from round 4 onward. The second row shows that the average return in B is significantly higher than in CR in all rounds but the first. Meanwhile, the third row shows that return in FP is significantly higher than in CR at the beginning of a supergame (p< 0.05), but this difference becomes insignificant from the third round onwards.[24]

Round:	1	2	3	4	5	6	All
FP Dummy	1.201	-0.0287	-0.783*	-2.109***	-2.472***	-1.863***	-0.886
(vs B)	(0.982)	(0.661)	(0.406)	(0.587)	(0.548)	(0.334)	(0.584)
	n=219	n=185	n=136	n=115	n=93	n=87	n=1196
B Dummy	1.749	2.808*	2.462*	2.900**	2.904**	2.514**	2.535**
(vs CR)	(0.993)	(1.285)	(1.075)	(0.925)	(0.943)	(1.090)	(1.020)
	n=223	n=189	n=167	n=153	n=132	n=118	n=1423
FP Dummy	2.950**	2.779*	1.679	0.790	0.432	0.652	1.649
(vs CR)	(1.096)	(1.189)	(1.044)	(1.050)	(0.954)	(1.086)	(0.989)
	n=220	n=178	n=137	n=116	n=115	n=95	n=1271

Standard errors clustered at the session level are reported in parentheses.
*** p< 0.01, ** p< 0.05, * p< 0.1

Table 2: Coefficients on treatment dummies in 21 round specific OLS regressions of return on treatment dummies. Pairwise treatment comparisons.

To confirm the impression that cooperation exhibits different within supergame dynamics across our treatments, we run a difference-in-difference OLS regression of return on a treatment dummy, a dummy for round 6 (against round 1), and a treatment-round interaction dummy, for each of the three possible treatment pairs (reported in Table 7 in the Appendix). This shows a larger increase in return between round 1 and 6 for B (p< 0.05) and CR (p< 0.1) compared to FP, but no difference in increase between B and CR. The regressions also show that the return increased significantly in B and CR (p< 0.01), but not in FP (Wald test $p = 0.871$).

The above results appear consistent with the importance of proposers sorting between cooperative and uncooperative responders. To the extent that proposers replaced uncooperative partners for (possibly) cooperative ones, sorting can account for the finding

[24]The return averaged across all rounds of the supergame is 6.9 in B, 6.0 in FP and 4.4 in CR; only the difference between B and CR is significant (p< 0.05). Because of inter-round dynamics, we do not believe that this comparison is particularly meaningful. The average is affected by the randomly determined length of the last four last supergames, and differences in the ability to punish and sort both predict a larger average return in B than in FP than in CR.

that cooperation rates in B and FP were initially similar, but increased only in B. Further evidence for the importance of sorting comes from the comparison between CR and FP. The finding that there is minimal difference in cooperation rates, after a few rounds of the supergame, is remarkable. It is consistent even with the extreme assumption of behavioral responder types, and as such, suggests that incentives for responders created by the threat of punishment were of limited importance in later rounds of a supergame. As outlined in the theory section, lower cooperation in CR than in FP at the start of the supergame can be explained by sorting, to the extent that there were greater incentives for (and greater use of) starting small strategies.

An implication of the extreme assumption of behavioral responders is that there should be little difference in return levels in round 1 of the supergame after controlling for the size of a responder's transfer. Table 8 in the Appendix reports the results of OLS regressions of return on treatment dummies and the transfer in pairwise treatment comparisons for round 1 data.[25] This suggests that after controlling for the transfer returns were marginally significantly higher in B than FP ($p < 0.1$) but CR was not statistically different from the other treatments. We interpret this result as failing to offer a convincing rejection of even this extreme assumption, and thus as evidence suggesting that the threat of punishment played a limited role in generating cooperation.

How robust are the results of Table 2 to alternative ways of analyzing the data? In the Online Appendix we present two alternatives: in Table 3 of the Online Appendix we first average return data at the level of a proposer in a round, before running the same 21 regressions (this gives one observation per proposer per regression); in Table 4 of the Online Appendix we pool data from all three treatments and run 7 OLS regressions of return on a treatment B dummy and a CR dummy. The exact levels of significance of treatment differences vary slightly, but both show that the average return in B and FP was not significantly different in round 1 and 2, but was significantly larger in B from round 3 onwards. Both similarly show that the average return was significantly larger in FP than CR in round 1 and 2 but was not significantly different in later rounds.[26]

4.2 Sorting through replacement

We hypothesized that cooperation in treatment B would be higher than in FP in later rounds of the supergame because proposers had greater opportunities for sorting by replacing uncooperative responders. The previous subsection illustrated aggregate cooperation data consistent with that prediction. This subsection examines more closely

[25] We do not make this comparison for later rounds because selection issues may bias the results. In particular, on finding an uncooperative responder in B a proposer may replace her in round 2 and make a large transfer to a new responder. In FP that same proposer might reduce her transfer to zero. In CR, the proposer has no option to reduce her transfer.

[26] Our results are also robust to using Tobit regressions to account for the fact that the return is "censored" above by ten and below by zero.

whether the observed pattern is consistent with the hypothesized mechanism.

For sorting to work, shirking (returning less than requested) should positively predict replacement in B. This is indeed the case. After the first round of a relationship (defined as the first time a proposer made a positive transfer to a particular responder) 78.7% of responders who shirked were replaced, compared to only 8% of those who complied with the request. A probit regression of replacement on shirking is significant at $p < 0.01$. Interestingly, repeated game punishment theory does not require proposers to punish shirking workers by replacing them. Indeed, proposers could punish equally effectively (with no loss of efficiency) without replacement, by reducing transfers to zero for one round and requesting a high return. However, no proposer ever actually did this in B.

In the simple sorting model with behavioral responder types, if a relationship survived for more than one round it should be completely stable thereafter (uncooperative types are sorted out immediately). Even under a richer specification of responder types, if sorting is important, we should expect relationships to become much more stable conditional on their initial survival. Figure 2 illustrates the extent to which this was true.

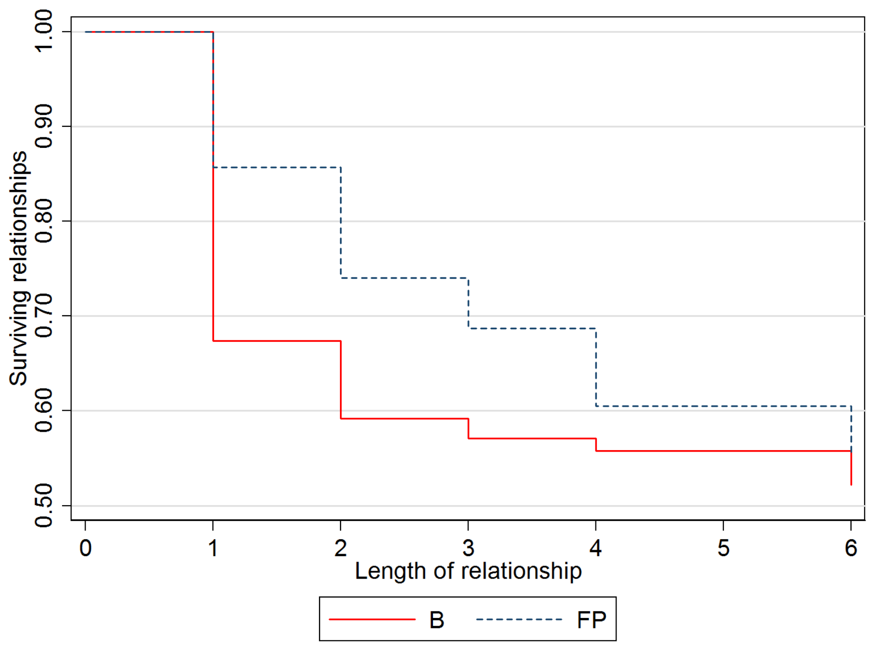

Figure 2: Relationship survival curves (Kaplan-Meier estimates). Failure is a round of zero transfer.

Consistent with the above definition of the start of a proposer–responder relationship, we define relationship failure as the first time that responder did not receive a positive transfer. Relationships do not fail because of the termination of the supergame; we treat that as censored data. For a relationship to be in its sixth round, the proposer must have first made a positive transfer to the responder 6 rounds previously, and have made positive transfers to her in all rounds since. Over 30% of relationships in B failed after the first round under these definitions, however, conditional on surviving to a second round the hazard rate was much lower and decreased further over time. Confirming this impression of increased stability later in a relationship, a dummy for first round of the relationship predicts its failure in a probit regression ($p < 0.01$). Notice, that according to our definition, we cannot meaningfully to talk about relationship failure in CR (as transfers could not be lowered).

Under the extreme assumption of behavioral responder types, proposers should adopt grim trigger strategies in FP. The optimality of such a strategy is less clear with a richer set of "somewhat uncooperative" responder types, however. The survival graph suggests that proposers were less likely to terminate a relationship immediately in FP compared to B, although strikingly, almost exactly the same fraction of many relationships had failed by round 6. In FP (as with B), shirking in the first round of a relationship predicts its immediate failure ($p < 0.01$), and the conditional likelihood of relationship failure is again predicted by a first round dummy ($p < 0.01$).

The key difference between the treatments is what happened after relationships failed. In B a new relationship could emerge with another responder, but this was impossible in FP. One way to compare the fraction of functioning relationships is to consider the fraction of proposers who made a positive transfer. Sorting predicts that these should be same at the start of the supergame but lower in FP than B in later rounds.

The left panel of Figure 3 shows that the fraction of proposers making positive transfers rose by more than 8 percentage points between rounds 1 and 6 of the supergame in B but fell by 14 percentage points in FP over the same period.[27,28] The first row of Table 9 in the Appendix reports coefficients (and standard errors) on a FP dummy in round specific probit regressions of a positive transfer dummy on an FP treatment dummy. Positive transfers were significantly more frequent in B than FP in round 4 ($p<0.1$), and round 6 ($p<0.05$). Confirming this impression of an increasing difference between the treatments, the first column of Table 10 in the Appendix reports two difference-in-difference probit regressions of a positive transfer dummy on an FP dummy, a round 6 dummy and an interaction term (on data for round 1 and 6). The negative coefficient on the interaction term is significant ($p < 0.01$), showing that the fraction of functioning relationships fell further in FP than in B.

An alternative way to compare the fraction of functioning relationships is to consider the fraction of proposers who received a positive return. The lower panel of Figure 3 shows that this rose by 8 percentage points between round 1 and 6 in B and fell by

[27]16 out of 111 (14.4%) round 1 offers were associated with a zero transfer in treatment B. Only 3 responders facing such offers made a positive return.

[28]The fact that more than 80% of transfers were positive in round 6 in FP despite almost 40% of relationships failing before their sixth round (a period of zero transfer, see Figure 2), shows that many proposers did not adopt grim trigger strategies.

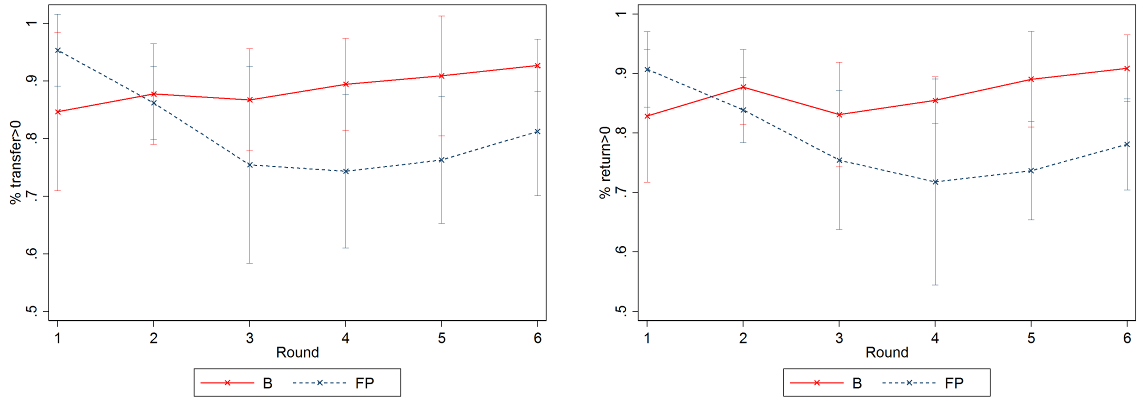

Figure 3: Shares of proposers making positive transfers $t > 0$ (upper panel) and responders making positive returns $r > 0$ (lower panel) by round.[29]

almost 13 percentage points in FP. The second row of Table 9 in the Appendix shows that significantly more relationships had positive return in B than in FP after round 4 ($p < 0.1$), rounds 5 and 6 ($p < 0.05$), and for all rounds ($p < 0.01$) of the supergame. Another difference-difference specification (second column, Table 10) reveals that the fraction of positive returns increased significantly more between rounds 1 and 6 in B compared to FP (interaction term $p < 0.01$). These results are consistent with the prediction of sorting through replacement.

4.3 Sorting through starting small

Central to our sorting explanation of aggregate cooperation patterns is the claim that proposers had greater incentives to start small in CR. Nonetheless, starting small could be profitable in all treatments. Table 3 suggests that many proposers did indeed choose to start small, and that this was more common in CR. In all three treatments more than half of all proposers started small in the sense of requesting less than the maximum return (of 10) in the first round of a relationship (as defined in section 4.2), with this most common in CR where more than 90% demanded less than the maximum. Table 11 in the Appendix reveals that this fraction in CR was significantly larger than in B ($p < 0.05$) or FP ($p < 0.01$); there is no significant difference between treatment B and FP. An OLS regression of initial requested return on treatment dummies further confirms this impression, with CR less than B ($p < 0.1$) and FP ($p < 0.05$), but no significant difference between B and FP.

Sorting logic suggests that proposers should promote cooperative responders (increasing transfers and the requested return), but not uncooperative types. This can be broken

[29] 95% confidence intervals from round specific probit regressions on treatment dummies (clustering at session level).

	Request	% Request< 10
B	7.025 (8.0)	68.2
FP	7.858 (8.5)	51.8
CR	4.491 (4.0)	90.9

Table 3: Average requested return (median in brackets) and percentages of sub-maximal requests in first round of a relationship

B	$\Delta\tilde{r} \leq 0$	$\Delta\tilde{r} > 0$
$\Delta t \leq 0$	19.35	6.45
$\Delta t > 0$	9.68	64.52

FP	$\Delta\tilde{r} \leq 0$	$\Delta\tilde{r} > 0$
$\Delta t \leq 0$	29.55	4.55
$\Delta t > 0$	2.27	63.64

CR	$\Delta\tilde{r} \leq 0$	$\Delta\tilde{r} > 0$
$\Delta t \leq 0$	35.80	12.35
$\Delta t > 0$	9.88	41.98

Table 4: Change of transfer and requested return from first to second round of a relationship (in %) if initial requested return<10

down into two predictions: firstly, transfers and requests should increase in lockstep (our definition of promotion), and secondly, shirking should negatively predict promotion.

Table 4 shows that transfers and requested returns did indeed move together in all three treatments. Off-diagonal elements where either transfers or requested return increased alone account for less than 22% of observations in all treatments. The table looks at proposer behavior in the second round of a relationship conditional on employment (a positive transfer) and an initial requested return below 10 (to allow for an increase). Confirming this interdependence, a probit regression of "increased return" on "increased transfer" is highly significant in all treatments (p< 0.01).

Table 5 confirms that promotion was tightly bound to responder performance. More than 60% of responders were promoted after providing the requested return in every treatment. After shirking, very few responders were promoted in either B or CR, although surprisingly as many as 45% were promoted in FP. The table is based on the first round of a relationship conditioning again on an initial requested return below 10. Table 12 in the Appendix reports a probit regression, showing that shirking significantly negatively affected the likelihood of promotion in B and CP (p< 0.01), but not in FP ($p = 0.107$).

Figure 4 plots average requested return against relationship length within *high-functioning* relationships, defined as relationships in which the responder had never shirked prior to that round. The figure again shows the special importance of starting small to the CR treatment. By the sixth round of these relationships, the average requested return in CR was 8.875, which is not significantly different from FP ($p = 0.894$) and only marginally

B	(Not)	Promoted
$r < \tilde{r}$	96.67	3.33
$r \geq \tilde{r}$	38.10	61.90

FP	(Not)	Promoted
$r < \tilde{r}$	54.55	45.45
$r \geq \tilde{r}$	35.71	64.29

CR	(Not)	Promoted
$r < \tilde{r}$	87.18	12.82
$r \geq \tilde{r}$	32.56	67.44

Table 5: Relation of promotion to shirking (return<requested) in the first round of a relationship if initial requested return<10 (Row %).

(p< 0.1) lower than B (reported in Table 13 in the Appendix).

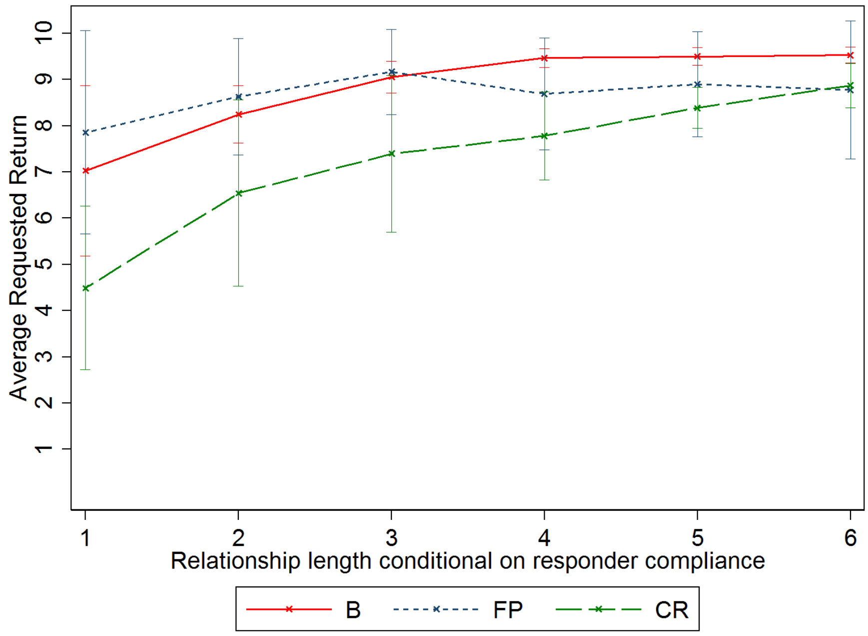

Figure 4: Requested return by relationship length, high-functioning relationships (i.e. conditional on responder compliance with the requested return in all previous rounds of relationship).[30]

If sorting worked along the lines of our simple model, at least within these high-functioning relationships there should be little difference in responder populations: they should all be cooperative types. To test for this, we regress return in round 6 of a high-functioning relationship on treatment dummies and the transfer. This mirrors our test of whether there was a different population of responders at the start of the supergame. We find no significant differences between treatments (reported in Table 13 in the Appendix).

It might seem that in addition to sorting, starting small can provide some incentive for responder cooperation, even if repeated game punishment theory tells us that this cannot work in equilibrium. One piece of evidence against this interpretation is that subjects start small even in the B and FP treatments, where responder incentives can be easily provided without increasing transfers. Moreover, a responder incentives explanation seems unable to explain starting small in terms of the requested return, or why transfers and requests move so closely together. If an increasing profile of transfers does provide incentives for responders, it would also seem consistent with high levels of return from the start of the game.

Nor do we believe starting small should be interpreted as subject learning (i.e. proposers realizing over time that higher transfers elicited higher returns). Evidence for this

[30] 95% confidence intervals are derived from regressions of return on treatment dummies in each round of a high-functioning relationship (clustering at the session level).

is provided by the fact that average proposer requests declined between the final round of one supergame and the first round of the next supergame in all treatments, with this significant for B ($p< 0.05$) and CR ($p< 0.1$). Moreover, an important part of our experimental design was to allow for learning about the supergame environment prior to data collection; recall that we use data on only the last four supergames.

BFF and FHM also find some transfer-return patterns consistent with starting small. Interestingly, however, BFF emphasize that high initial proposer transfers (starting big) is essential for relationship success, with the size of transfer in the first round of a relationship predicting relationship length. To test for this in our setting we run a Cox proportional-hazards regression for relationship survival (as defined in section 4.2)[31] on the transfer in the first round of the relationship for treatment B and FP separately (Table 14 in the Appendix). The coefficients on the transfer are not significant, suggesting no advantage to starting big.

4.4 Relationship building and surplus sharing

In this subsection, we replicate BFF's finding that surplus is shared evenly between the partners in successful long term relationships (in a treatment comparable to our treatment B). This is perhaps unsurprising given the literature on fairness concerns in experiments, however, notice that in our setup (with an indefinite as opposed to finite horizon) selfish agents did not need to pretend to be fair/reciprocative types to establish cooperation in equilibrium.

We define the surplus created as $3r$ and therefore the share of surplus received by a proposer to be $\frac{4r-t}{3r}$ if $r \in (\frac{t}{4}, t)$, 0 if $r \leq \frac{t}{4}$, and 1 if $r \geq t$. To find the share of surplus requested by a proposer replace r with $\tilde{r}$ in the above definition. We define a division as "roughly equal" if it gave each party between 45% and 55% of the surplus (transfers associated with such divisions are *at most* one point from an exactly equal split). Table 6 reports the percentages of roughly equal proposed and received surplus divisions in round 1 and 6 of a relationship (defined in section 4.2), and round 6 of a high-functioning relationship (defined in section 4.3).

	Round 1	Round 6	Round 6 High functioning
B	44.2 (32.5)	78.5 (77.5)	82.4 (79.4)
FP	66 (37.7)	89.5 (63.2)	100 (100)
CR	25.7 (18.2)	37.9 (31.1)	73.3 (73.3)

Table 6: % of surplus shares proposed (and received) that were roughly equal, in rounds 1 and 6 of a relationship and round 6 of high-functioning relationships.

[31] Recall that relationships cannot fail in this way in CR.

In treatments B and FP 44.2% and 66% respectively of proposers' requests are roughly equal even in the first round of a relationship, with slightly fewer roughly equal shares received, 32.5% and 37.7% respectively. By round 6 of a relationship, these numbers had risen to 78.5% and 89.5% of proposed divisions, and 77.5% and 63.2% of received divisions.

Comparing these numbers to CR is more difficult for a number of reasons. The first is integer constraints: with a requested return of one, a transfer of less than two proposes a share of above 66%, while a transfer of strictly more than two proposes a share below 33%, which means roughly equal shares were impossible. A similar conclusion applies to a requested return of three. Given the prevalence of starting small in CR it is perhaps no surprise that only 25.7% of initial proposals were roughly equal and only 18.2% of received divisions were roughly equal. Because transfers could not be reduced, relationships in CR could not fail (in particular following responder shirking), which means the 37.9% proposed and 31.1% received roughly equal shares in round 6 of a relationship are again difficult to compare to B and FP.

For this reason we also compare divisions within high-functioning relationships (as defined in subsection 4.3). 73.3% of proposed and received shares are roughly equal in CR, which is comparable to the incidence of roughly equal shares in B (82.4% proposed, 79.4% received) and FP (100% of shares proposed and received).[32]

4.5 Learning

Our main qualitative results are not substantively affected by the exact cutoff which we use to allow for learning. Indeed, in the Online Appendix (Table 1) we report regressions of return on treatment dummies comparable to those reported in Table 2 using data for all supergames. This shows that average return is not significantly different in B and FP in the first two rounds of the supergame, but is significantly higher in B in later rounds. Return is significantly higher in FP than CR in the first round of the supergame, but is not in later rounds. We also report (in Table 2 of the Online Appendix) the same regressions for data after subjects had played at least 20 total rounds. As one anonymous referee suggested, learning may be by round or time, rather than by supergame. Unsurprisingly, this alternative positive learning cutoff gives results that are much closer to those found using data from the last four supergames.

These findings, of course, do not imply that learning didn't matter. The LHS of Figure 5 plots the change in average return between the last four supergames and all previous

[32]Looking at actual per-period earnings instead of surplus shares (Table 5 of the Online Appendix), we find that proposers in treatment B earned significantly more than proposers in the two other treatments ($p < 0.01$ compared to both FP and CR). Proposers in treatment FP also earned significantly more than proposers in treatment CR ($p < 0.05$). Looking at responders' per-period earnings, on the other hand, we find no significant differences between treatments ($F-$test $p = 0.548$).

supergames for rounds 1 to 6.[33] The figure shows that learning about the game appears associated with: an increased return of between 1.5 points in round 1 to 3 points by round 6 in treatment B; an increased return of between 0.5 and 1.5 points in FP; and essentially no increased return in CR.

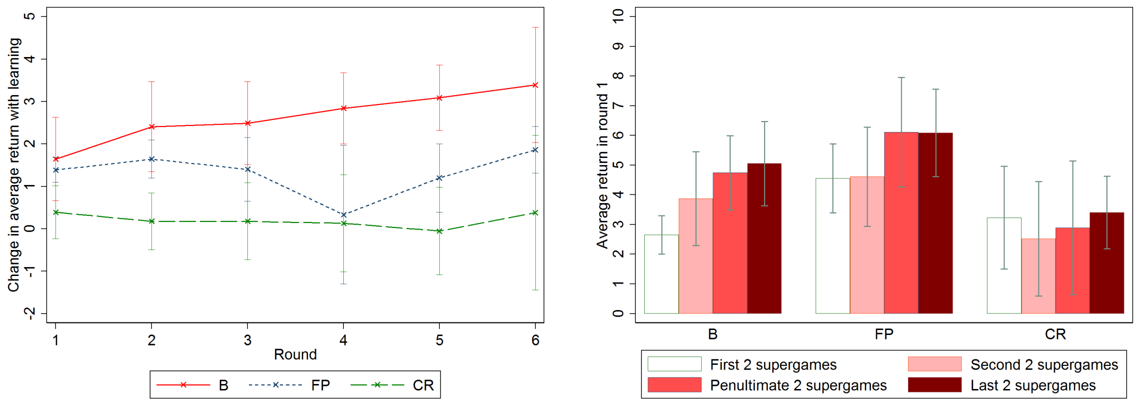

Figure 5: (LHS) Change of average return between last four supergames and previous supergames by round; (RHS): Average return in round one for pairs of supergames.

Table 15 confirms that changes of return differed by treatment. We regress return on treatment dummies, a dummy for the last four supergames and an interaction term for each pair of treatments for rounds 1-6 and for all rounds.[34] Results shows that average returns increase significantly more in B than FP, except for in rounds 1 and 2. Returns increased significantly more in B than CR regardless of round. Returns increased significantly more in FP than in CR except for in rounds 4 and 6.

As pointed out by an anonymous referee, in treatment B there were many more public offers made after round 1 than in FP and potentially this could have allowed for different amounts of learning (from others). To the extent that there was differential learning about (how to make and respond to) public offers, however, we should expect this to show up in differences of behavior in round 1, where offers were necessarily public. The fact that not only was the average return in B and FP was not significantly different in round 1 in the last 4 supergames in (Table 2) but that the increase in the round 1 return as a result of learning was not significantly different reassures us that our main results are most likely not affected by differences in the ability to learn from public offers.

Instead, the fact that returns increased more in B than FP only in later rounds of the supergame suggests that different behavior in the last four supergames was due to

[33]It also includes 95% confidence intervals for this change derived from 6 round specific regressions of return on treatment dummies, dummies for the last four supergames, and interaction terms (clustering at the session level).

[34]In order to report the results compactly, we reports only the coefficients (and standard errors) on interaction terms (our variable of interest), with row 1 comparing treatments FP and B, row 2 comparing B and CR, and row 3 comparing FP and CR.

responders learning how to effectively sort workers through replacement in treatment B (which was possible in B but not FP). Learning to sort through replacement should mean replacing uncooperative responders, and not replacing cooperative ones. To test for this, in Table 16 in the Appendix we report a probit regression of a dummy for whether the responder was replaced on a dummy for the last 4 supergames, a dummy for whether the responder shirked (return<requested) and an interaction term, for data in the first round of a supergame. A negative significant ($p<0.05$) coefficient on the last 4 supergames dummy shows that the probability of replacement of a cooperative responder was lower after learning, while the positive significant ($p<0.01$) coefficient on the interaction dummy shows that the probability of replacement of a uncooperative responder was higher after learning. Taken together, this suggests that proposers did indeed learn to use replacement to effectively sort responders.

The fact that return increased more in FP than CR, might seem to suggest that given even more time for learning (more repetitions of the supergame), the average return in FP would end up higher than in CR even for later rounds of the supergame. We cannot completely dismiss this possibility, however, we are skeptical of it. The RHS of Figure 5 attempts to show a more granular learning trajectory. It plots the average return in round 1 for the first two supergames, the next two supergames, the penultimate two supergames (i.e. when either 3 or 4 supergames remain) and the final two supergames.[35] The key thing to notice is that for all three treatments there is little increase in the average return between the penultimate two supergames and the last two supergames, indeed the average return actually decreases slightly in FP. This suggests that further repetitions of the supergame would be unlikely to affect our main conclusions.

One reason that can partially explain the lower increase of return in CR compared to FP that is consistent with our simple sorting model, is proposers learning about the incentives to start small in CR. In column 5 (and 6) of Table 17 we run an OLS regression (probit regresssion) of requested return (a dummy for requested return<10) on a dummy for the last four supergames, an FP dummy, and an interaction term for round 1 data from CR and FP. The positive significant ($p<0.05$ and $p<0.01$) coefficients on the interaction terms suggests that proposers did indeed learn about the larger incentives to start small in CR.[36] Columns 1-4 of Table 17 report other treatment comparisons. Notably, the interaction terms are not significant when comparing B and FP.

5. Conclusion

Beyond the small literature on starting small, the economics profession has focused much less attention on sorting in repeated games, than on incentives created by threats

[35] Comparing the round 1 return gives us data from all the supergames, while we have less data for later rounds.

[36] This change is of course also consistent with repeated game punishment theory.

of future punishment. Existing experimental studies on repeated gift-exchange games have linked lower cooperation when subjects cannot identify previous partners or face dismissal barriers to responders' (lack of) incentive to cooperate without the threat of punishment. They do not consider a sorting explanation. Our experiment independently varied a proposer's ability to replace a responder (her ability to sort) and her ability to punish. Remarkably, we found that the availability of punishments does little to explain cooperation patterns. By contrast, a simple sorting model, with an extreme assumption that all responders are behavioral cooperative or uncooperative types, explains our data remarkably well.

Sorting, as we consider it, is about finding a trustworthy partner, rather than providing incentives for an inherently self-interested one. Our results suggests that the ability to sort may be critical for explaining why and when people cooperate in repeated settings. It is not that we regard incentives per se as unimportant (recall that repeated interaction can increase cooperation in our simple sorting model precisely because it improves proposers' incentives to take the risky step of starting a relationship). However, the broader implications of a sorting model are quite different from a repeated game punishment theory model. For example, the inefficiency of real world institutions, such as employment protection legislation, may be much more affected by improvements in candidate selection than by instituting performance related pay. We certainly believe more work should be done to establish the role of sorting more clearly in such contexts.

References

Aimone, J. A., Iannaccone, L. R., Makowsky, M. D. & Rubin, J. (2013), 'Endogenous group formation via unproductive costs', *The Review of economic studies* **80**(4), 1215–1236.

Andreoni, J., Kuhn, M. A. & Samuelson, L. (2016), 'Starting small: Endogenous stakes and rational cooperation', *Working Paper* .

Andreoni, J. & Samuelson, L. (2006), 'Learning to start small', *Journal of Economic Theory* **127**, 117–154.

Brown, M., Falk, A. & Fehr, E. (2004), 'Relational contracts and the nature of market interactions', *Econometrica* **72**(3), 747–780.

Castillo, M. & Petrie, R. (2010), 'Discrimination in the lab: Does information trump appearance?', *Games and Economic Behavior* **68**(1), 50–59.

Charness, G., Cobo-Reyes, R. & Jiméne, N. (2014), 'Identities, selection, and contributions in a public-goods game', *Games and Economic Behavior* **87**, 322–338.

Charness, G. & Rabin, M. (2002), 'Understanding social preferences with simple tests', *The Quarterly Journal of Economics* **117**(3), 817–869.

Charness, G. & Yang, C.-L. (2014), 'Starting small toward voluntary formation of efficient large groups in public goods provision', *Journal of Economic Behavior Organization* **102**, 119–132.

Coricelli, G., Fehr, D. & Fellner, G. (2004), 'Partner selection in public goods experiments', *Journal of Conflict Resolution* **48**(3), 356–378.

Croson, R., Fatas, E., Neugebauer, T. & Morales, A. J. (2015), 'Excludability: A laboratory study on forced ranking in team production', *Journal of Economic Behavior Organization* **114**, 13–26.

Dal Bó, P. (2005), 'Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games', *American Economic Review* **95**(5), 1591–1604.

Dal Bó, P. & Fréchette, G. (2011), 'The evolution of cooperation in infinitely repeated games: Experimental evidence', *American Economic Review* **101**(1), 411–429.

Embrey, M., Fréchette, G. & Yuksel, S. (2016), Cooperation in the finitely repeated prisioner's dilemma. Working Paper.

Engle-Warnick, J. & Slonim, R. L. (2004), 'The evolution of strategies in a repeated trust game', *Journal of Economic Behavior & Organization* **55**(4), 553–573.

Eriksson, T. & Villeval, M. C. (2012), 'Respect and relational contracts', *Journal of Economic Behavior & Organization* **81**(1), 286–298.

Falk, A., Huffman, D. & MacLeod, W. B. (2008), 'Institutions and contract enforcement', *Working Paper* **NBER 13961**.

Fehr, E., Kirchler, E., Weichbold, A. & Gächter, S. (1998), 'When social norms overpower competition: Gift exchange in experimental labor markets', *Journal of Labor Economics* **16**(2), 324–351.

Fehr, E., Kirchsteiger, G. & Riedl, A. (1993), 'Does fairness prevent market clearing? an experimental investigation', *Quarterly Journal of Economics* **108**(2), 437–459.

Fischbacher, U. (2007), 'z-tree: Zurich toolbox for ready-made economic experiments', *Experimental Economics* **10**(2), 171–178.

Gächter, S. & Thöni, C. (2005), 'Social learnign and voluntary cooperation among like-minded people', *Journal of the European Economic Association* **3**, 303–314.

Grimm, V. & Mengel, F. (2009), 'Cooperation in viscous populations?experimental evidence', *Games and Economic Behavior* **66**(1), 202–220.

Hauk, E. & Nagel, R. (2001), 'Choice of partners in multiple two-person prisoner's dilemma games an experimental study', *Journal of conflict resolution* **45**(6), 770–793.

Janssen, M. A. (2008), ‘Evolution of cooperation in a one-shot prisoner’s dilemma based on recognition of trustworthy and untrustworthy agents’, *Journal of Economic Behavior & Organization* **65**(3), 458–471.

Kurzban, R., McCabe, K., Smith, V. L. & Wilson, B. J. (2001), ‘Incremental commitment and reciprocity in a real-time public goods game’, *Personality and Social Psychology Bulletin* **27**(12), 1662–1673.

Kurzban, R., Rigdon, M. L. & Wilson, B. J. (2008), ‘Incremental approaches to establishing trust’, *Experimental Economics* **11**(4), 370–389.

Lazear, E. P. (2000), ‘Performance pay and productivity’, *The American Economic Review* **90**(5), 1346–1361.

Lazear, E. P., Malmendier, U. & Weber, R. A. (2012), ‘Sorting in experiments with application to social preferences’, *American Economic Journal: Applied Economics* **4**(1), 136–163.

MacLeod, W. B. & Malcolmson, J. (1989), ‘Implicit contracts, incentive compatibility, and involuntary unemployment’, *Econometrica* **57**(2), 447–480.

Murphy, A. H. & Winkler, R. L. (1970), ‘Scoring rules in probability assessment and evaluation’, *Acta Psychologica* **34**, 273–286.

Page, T., Putterman, L. & Unel, B. (2005), ‘Voluntary association in public goods experiments: reciprocity, mimicry and efficiency’, *The Economic Journal* **115**(506), 1032–1053.

Rigdon, M. L., McCabe, K. A. & Smith, V. L. (2007), ‘Sustaining cooperation in trust games’, *The Economic Journal* **117**(522), 991–1007.

Roe, B. E. & Wu, S. Y. (2009), ‘Do the selfish mimic cooperators? experimental evidence from finitely-repeated labor markets’, *Experimental Evidence from Finitely-Repeated Labor Markets* .

Watson, J. (1999), ‘Starting small and renegotiation’, *Journal of Economic Theory* **85**, 52–90.

6. Appendix: regression results

We report cluster-robust standard errors in parentheses, *** p< 0.01, ** p< 0.05, * p< 0.1.

	B vs. FP	B vs. CR	FP vs. CR
	(1)	(2)	(3)
FP Dummy	1.201		2.950**
	(.985)		(1.099)
B Dummy		1.749	
		(.995)	
Round 6 Dummy	2.899***	2.134***	2.134***
	(.503)	(.349)	(.350)
Round 6 X FP	-3.063**		-2.298*
	(1.100)		(1.038)
Round 6 X B		.765	
		(.612)	
Const.	4.901***	3.152***	3.152***
	(.614)	(.783)	(.784)
Obs.	306	341	315
F statistic	36.45	90.107	55.097
R^2	.078	.17	.128

Table 7: OLS regressions of return on treatment dummies, round 6 dummy and treatment-round interaction dummies. Data from round 1 and 6, pairwise treatment comparisons.

	B vs. FP	B vs. CR	FP vs. CR
	(1)	(2)	(3)
FP Dummy	-.525*		-.126
	(.249)		(.348)
B Dummy		.241	
		(.285)	
Transfer	.325***	.337***	.314***
	(.013)	(.016)	(.022)
Const.	.644***	.255	.447
	(.177)	(.281)	(.261)
Obs.	219	223	220
F statistic	304.946	227.562	149.936
R^2	.657	.701	.628

Table 8: OLS regressions of (responder) return on treatment dummies and (proposer) transfer. Data from round 1, pairwise treatment comparisons.

Round:	1	2	3	4	5	6	All
			Positive transfer ($t > 0$)				
FP Dummy	1.315	-0.137	-0.755	-1.075*	-1.133	-1.079**	-0.493
(vs B)	(0.899)	(0.498)	(0.613)	(0.560)	(0.714)	(0.508)	(0.465)
			Positive return ($r > 0$)				
FP Dummy	0.705	-0.318	-0.471	-0.842*	-1.070**	-1.030**	-0.501*
(vs B)	(0.556)	(0.365)	(0.454)	(0.466)	(0.476)	(0.416)	(0.298)
	n=219	n=185	n=136	n=115	n=93	n=87	n=1196

Table 9: Coefficients on FP dummy in round specific probit regressions of positive transfer dummy (row 1) and positive return dummy (row 2) on FP dummy. FP and B data.

	$t > 0$	$r > 0$
	(1)	(2)
FP Dummy	.659	.375
	(.441)	(.297)
Round 6 Dummy	.433***	.386***
	(.157)	(.121)
Round 6 X FP	-1.228***	-.934***
	(.444)	(.306)
Const.	1.023***	.950***
	(.296)	(.224)
Obs.	306	306
χ^2 statistic	20.022	18.724
Pseudo-R^2	.048	.024

Table 10: Probit regression of positive transfer dummy (column (1)) and positive return dummy (column (2)) on FP dummy, round 6 dummy and interaction dummy. Data from round 1 of a relationship.

	B vs. FP		B vs. CR		FP vs. CR	
	(1)	(2)	(3)	(4)	(5)	(6)
FP Dummy	-.425	.833			-1.288***	3.368**
	(.384)	(1.337)			(.458)	(1.314)
B Dummy			-.863**	2.535*		
			(.339)	(1.189)		
Const.	.472***	7.025***	1.335***	4.491***	1.335***	4.491***
	(.163)	(.860)	(.298)	(.822)	(.298)	(.823)
Obs.	263	263	267	267	216	216
F (χ^2) statistic	1.222	.388	6.47	4.54	7.906	6.571
(Pseudo-)R^2	.02	.02	.074	.152	.168	.282

Table 11: Odd columns: Probit regression of dummy for requested return<10 on treatment dummies. Even columns: OLS regression of requested return on treatment dummies. All columns: Data from round 1 of a relationship, pairwise treatment comparisons.

	B	FP	CR
	(1)	(2)	(3)
Shirked Dummy	-2.137***	-.480	-1.587***
	(.537)	(.298)	(.527)
Const.	.303	.366	.452*
	(.278)	(.269)	(.261)
Obs.	93	50	82
χ^2 statistic	15.857	2.592	9.083
Pseudo-R^2	.272	.026	.244

Table 12: Treatment specific probit regressions of promotion dummy (increased request and transfer) on shirking dummy (return<requested). Data from round 1 of a relationship if initial request<10.

	B vs. FP		B vs. CR		FP vs. CR	
	(1)	(2)	(3)	(4)	(5)	(6)
FP Dummy	-.752*	-.221			-.097	.074
	(.380)	(.281)			(.723)	(.192)
B Dummy			.654*	.297		
			(.351)	(.233)		
Transfer		.320***		.339***		.359***
		(.040)		(.023)		(.014)
Const.	9.529***	1.987**	8.875***	1.257**	8.875***	.838***
	(.174)	(.946)	(.289)	(.506)	(.434)	(.313)
Obs.	43	43	50	50	25	25
F statistic	3.922	33.973	3.483	123.83	.018	327.687
R^2	.087	.629	.068	.84	.0008	.968

Table 13: Odd columns: OLS regressions of requested returns on treatment dummies. Even columns: OLS regressions of actual (responder) returns on treatment dummies and proposer transfer. All columns: Data from round 6 of high-functioning relationships, pairwise treatment comparisons.

	B	FP
	(1)	(2)
Transfer	-.010	.022
	(.007)	(.016)
Obs.	449	298
χ^2 statistic	1.92	1.93
Pseudo-R^2	.0006	.004

Table 14: Treatment specific cox proportional hazard regressions for relationship survival on transfer in round 1 of a relationship.

Round:	1	2	3	4	5	6	All
FP Dummy × Last 4 Dummy	-0.258	-0.764	-1.088*	-2.511**	-1.892***	-1.528*	-1.252**
(vs B)	(0.477)	(0.536)	(0.572)	(0.854)	(0.520)	(0.683)	(0.472)
	n=518	n=421	n=298	n=244	n=201	n=163	n=2525
B Dummy × Last 4 Dummy	1.256*	2.232***	2.311***	2.714***	3.143***	3.011**	2.299***
(vs CR)	(0.541)	(0.584)	(0.621)	(0.659)	(0.597)	(1.055)	(0.430)
	n=515	n=424	n=333	n=292	n=264	n=230	n=2769
FP Dummy × Last 4 Dummy	0.998**	1.468***	1.224*	0.203	1.251*	1.482	1.047***
(vs CR)	(0.320)	(0.376)	(0.548)	(0.927)	(0.607)	(0.885)	(0.290)
	n=521	n=431	n=315	n=260	n=245	n=185	n=2654

Table 15: Coefficients on interaction dummies in 21 round specific OLS regressions of return on treatment dummies, a dummy for the last 4 supergames and interaction dummies. Pairwise treatment comparisons.

	Replaced (B)
Last 4 Dummy	-0.733**
	(0.328)
Shirked Dummy	1.327***
	(0.203)
Last 4 Dummy×Shirked Dummy	1.091***
	(0.423)
Constant	-0.776***
	(0.227)
Obs.	170
χ^2 statistic	.
Pseudo-R^2	0.343

Table 16: Probit regression of a dummy for replacement on a dummy for the last 4 supergames, a dummy for shirking (return<requested) and an interaction dummy. Data from round 1, treatment B.

	B vs. FP		B vs. CR		FP vs. CR	
	(1)	(2)	(3)	(4)	(5)	(6)
Last 4 Dummy	0.0159	-0.333***	-1.083	0.330	-1.083	0.330
	(0.464)	(0.102)	(0.606)	(0.233)	(0.606)	(0.233)
FP Dummy	0.515	-0.257			1.058	-0.324
	(1.083)	(0.262)			(0.989)	(0.260)
Last 4 Dummy×FP Dummy	0.993	-0.195			2.092**	-0.858***
	(0.532)	(0.177)			(0.659)	(0.274)
B Dummy			0.543	-0.0674		
			(0.775)	(0.138)		
Last 4 Dummy×B Dummy			1.099	-0.663***		
			(0.763)	(0.254)		
Constant	6.293***	0.858***	5.750***	0.925***	5.750***	0.925***
	(0.630)	(0.101)	(0.451)	(0.0949)	(0.451)	(0.0949)
Obs.	493	493	472	472	479	479
χ^2 statistic		24.76		14.87		18.08
Pseudo-R^2		0.0355		0.0289		0.0772
F statistic	5.565		1.184		6.357	
R^2	0.0369		0.0423		0.144	

Table 17: Odd columns: OLS regressions of requested return on a dummy for the last 4 supergames, a dummy treatment dummies and interaction dummies. Even columns: Probit regressions of dummy for requested return<10 on a dummy for the last 4 supergames, a dummy treatment dummies and interaction dummies. All columns: data from round 1, pairwise treatment comparisons.

Online Appendix for

Finding cooperators: Sorting through repeated interaction

Mark Bernard Jack Fanning Sevgi Yuksel

Tables and Figures

Round:	1	2	3	4	5	6	All
FP Dummy	1.318	0.263	-0.473**	-1.042**	-1.625**	-1.117*	-0.378
(vs B)	(0.841)	(0.534)	(0.195)	(0.318)	(0.543)	(0.504)	(0.524)
	n=518	n=421	n=298	n=244	n=201	n=163	n=2525
B Dummy	1.038	1.642	1.373	1.687	1.332	1.107	1.408
(vs CR)	(0.947)	(1.149)	(1.062)	(0.947)	(0.882)	(1.093)	(1.030)
	n=515	n=424	n=333	n=292	n=264	n=230	n=2769
FP Dummy	2.356*	1.905	0.900	0.645	-0.293	-0.0105	1.030
(vs CR)	(1.100)	(1.055)	(1.048)	(0.969)	(0.887)	(1.033)	(0.986)
	n=521	n=431	n=315	n=260	n=245	n=185	n=2654

Standard errors clustered at the session level are reported in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1. Coefficients on treatment dummies in 21 round specific OLS regressions of return on treatment dummies. Data from all supergames (not last 4 supergames), pairwise treatment comparisons.

Round:	1	2	3	4	5	6	All
FP Dummy	0.869	-0.182	-1.034**	-2.290***	-3.162***	-2.425**	-1.097*
(vs B)	(0.882)	(0.571)	(0.389)	(0.506)	(0.685)	(0.689)	(0.571)
	n=252	n=211	n=163	n=129	n=108	n=95	n=1442
B Dummy	1.650	2.726*	2.411*	2.874**	2.882**	2.719**	2.425*
(vs CR)	(1.018)	(1.225)	(1.074)	(0.949)	(0.926)	(1.085)	(1.051)
	n=244	n=196	n=173	n=174	n=160	n=139	n=1660
FP Dummy	2.519*	2.544*	1.376	0.585	-0.280	0.294	1.327
(vs CR)	(1.144)	(1.134)	(1.077)	(1.019)	(1.029)	(1.251)	(1.014)
	n=274	n=225	n=158	n=137	n=144	n=110	n=1560

Standard errors clustered at the session level are reported in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2. Coefficients on treatment dummies in 21 round specific OLS regressions of return on treatment dummies. Data after subjects had played at least 20 total rounds (not last 4 supergames), pairwise treatment comparisons.

Round:	1	2	3	4	5	6	All
FP Dummy	1.158	0.252	-0.852*	-2.232**	-2.401***	-1.877***	-1.053*
(vs B)	(0.979)	(0.708)	(0.397)	(0.791)	(0.647)	(0.269)	(0.467)
	n= 55	n= 55	n= 55	n= 55	n= 53	n= 53	n= 366
B Dummy	1.792	2.646*	2.717*	3.071**	3.020**	3.305**	2.716**
(vs CR)	(0.990)	(1.369)	(1.246)	(0.990)	(1.046)	(1.377)	(1.095)
	n= 56	n= 56	n= 56	n= 56	n= 55	n= 55	n= 383
FP Dummy	2.950**	2.898*	1.865	0.839	0.619	1.428	1.664
(vs CR)	(1.104)	(1.308)	(1.209)	(1.213)	(1.154)	(1.373)	(1.098)
	n= 55	n= 55	n= 55	n= 55	n= 54	n= 54	n= 375

Standard errors clustered at the session level are reported in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3. Coefficients on treatment dummies in 21 round specific OLS regressions of return on treatment dummies. Data first averaged by proposer and round, pairwise treatment comparisons.

Round:	1	2	3	4	5	6	All
CR Dummy	-2.356*	-1.905*	-0.900	-0.645	0.293	0.0105	-1.030
	(1.075)	(1.031)	(1.025)	(0.947)	(0.867)	(1.010)	(0.963)
B Dummy	-1.318	-0.263	0.473**	1.042***	1.625**	1.117**	0.378
	(0.822)	(0.522)	(0.191)	(0.311)	(0.530)	(0.492)	(0.512)
Constant	5.286***	5.495***	5.129***	4.943***	4.538***	5.085***	5.251***
	(0.698)	(0.194)	(0.0640)	(0.261)	(0.380)	(0.244)	(0.297)
Obs	777	638	473	398	355	289	3974
F statistic	2.514	1.778	3.507	6.723	4.914	2.610	1.018
R^2	0.0766	0.0519	0.0237	0.0353	0.0304	0.0178	0.0250

Standard errors clustered at the session level are reported in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4. Coefficients on treatment dummies in 7 round specific OLS regressions of return on treatment dummies.

	B vs. FP	B vs. CR	FP vs. CR	B vs. FP	B vs. CR	FP vs. CR
	(1)	(2)	(3)	(4)	(5)	(6)
B Dummy	2.528***	5.639***		.131	1.966	
	(.684)	(1.274)		(1.185)	(1.879)	
FP Dummy			3.111**			1.835
			(1.309)			(1.766)
Const.	7.534***	4.423***	4.423***	10.517***	8.682***	8.682***
	(.528)	(1.197)	(1.197)	(.705)	(1.620)	(1.620)
Obs.	1196	1423	1271	1196	1423	1271
F statistic	13.662	19.607	5.653	.012	1.095	1.079
R^2	.023	.099	.029	.0001	.027	.021

Table 5. Coefficients on treatment dummies in OLS regressions of per-period earnings. Columns (1)-(3) test proposer earnings, columns (4)-(6) responder earnings.

Sample Instructions

INSTRUCTIONS

Welcome

You are about to participate in an experiment on decision-making. What you earn depends partly on your decisions, partly on the decisions of others, and partly on chance. Please turn off cell phones and similar devices now. Please do not talk or in any way try to communicate with other participants.

After reading the instructions, we will lead you through a tutorial to get you accustomed to the computer interface. If you have any questions after these, raise your hand and your question will be answered so everyone can hear.

During the experiment you will have the chance to earn experimental points. All points earned will be exchanged into Dollars. The exchange rate will be:

50 points = 1 dollar

You will receive the sum of money that you earned during the experiment in addition to your $ 10 initial show-up fee.

Prior to the experiment the 20 participants were divided into 2 groups: Proposers and Responders. There are 7 Proposers and 13 Responders. Your role will be fixed in this experiment. That is, if you are a Responder/Proposer, you will be a Responder/Proposer throughout the experiment.

General Instructions

In this experiment you will be asked to make decisions in several games. Each game consists of a sequence of rounds.

The length of each game is randomly determined. After each round, there is an 80% probability that the game will continue for at least another round. Specifically, after each round, whether the game continues for another round will be determined by a random number between 1 and 100 generated by the computer. If the number is lower than or equal to 80 the game will continue for at least another round, otherwise it will end. For example, if you are in round 2, the probability that there will be a third round is 80% and if you are in round 9, the probability that there will be a tenth round is also 80%.

At the beginning of each game, all participants will receive an identification number, which they will keep throughout that game. This will allow you to keep track of who you are playing with during a game. However, these identification numbers will be changed once a game ends, before a new game begins. This means that you will not be able to identify who you've played with in previous games. You should treat each game as independent.

Overview of a game

Each round of each game consists of three stages. In Stage 1 Proposers make offers to Responders; in Stage 2 Responders decide on which offers to accept. In Stage 3 Responders, who accepted offers, choose how much to return to the Proposers.

A round of each game

STAGE 1

- Every Proposer and Responder will be given an initial 10 points.
- Each Proposer decides whether or not to make an offer.
- An offer consists of two numbers: a transfer and a request. The transfer (a whole number between 0 and 40) specifies how many points will be transferred to a Responder who accepts the offer. The request (a whole number between 0 and 10) specifies how much the Proposer would like in return from the Responder.
- Each Proposer who decides to make an offer also decides if her offer is public or private. Now we explain the difference between these two kinds of offers.
    - If she chooses to make a **public offer**, it will be observed by all Responders. Additionally, if she wants, the Proposer can block any number of Responders from accepting the offer. These Responders will observe the offer, but will not be able to accept it. Responders will not see if other Responders are blocked from the offer or not.
    - If the Proposer chooses to make a **private offer**, only the Responder who accepted her offer in the last round will observe this offer. This means that it is possible for the Proposer to make a private offer only if her offer was accepted in the previous round. Therefore, private offers are not available in the first round of a game.

STAGE 2

- Each Responder observes both private and public offers available to her.
- If she has a private offer, her first step is to decide whether to accept or reject it. If she accepts, she moves on to Stage 3. She can reject the private offer by clicking "Reject private offer". If the Responder doesn't specify a decision within the time limit, the program assumes she accepts the private offer.
- If a private offer wasn't made, or was rejected, the Responder proceeds to select from public offers available to her. She can also always decide not to accept any offer at all.
- Note that since public offers can be accepted by other Responders they might run out.

STAGE 3

- Responders who have accepted offers decide how much to return to the Proposer (a whole number between 0 and 10). Remember that the Responder is not required to return the amount requested by the Proposer.
- Once Responders have chosen return amounts, individual payoffs are determined.
- If a Proposer does not submit an offer in the time allocated for Stage 1, or her offer is not accepted, she keeps her 10 points and waits for the next round.
- If a Responder does not have available offers, or chooses not to accept any, or doesn't specify a return amount in the time allocated for Stage 3 she keeps her 10 points and waits for the next round.

Payoffs

Points gained in each round are calculated as follows:

- *If an offer was made and accepted:*
  - ⇒ Proposer payoff = 10 (initial points) - transfer + 4 × return
  - ⇒ Responder payoff = 10 (initial points) + transfer - return
- *If an offer wasn't made or accepted:*
  - ⇒ Proposer payoff = 10 (initial points)
  - ⇒ Responder payoff = 10 (initial points)

The payoff table distributed with your instructions gives you an overview of how payoffs per round depend on transfer and return amounts.

Total payoffs for each game will be the sum of payoffs obtained from each round of that game. Total payoffs for the experiment will be the sum of payoffs for all games played, plus your show-up fee. Any losses will be deducted from your show-up fee.

Timing

In every round of every game:

- Stage 1 (Making offers) will last 30 seconds.
- Stage 2 (Accepting offers) will last up to 30 seconds.
- Stage 3 (Choosing returns) will last 15 seconds.
- The first 5 rounds will last 30 seconds longer to give you extra time to familiarize with the game.
- The first game to end after 70 minutes of play will be the last game

Summary

In each round of each game:

- Proposers make offers, either private or public, which specify a transfer amount and requested return.
- Proposers can only make private offers to Responders who accepted their offer in the last round. Public offers can be made available to any number of Responders.
- Responders decide to accept or reject their private offer (when available).
- If a private offer wasn't made, or was rejected, Responders choose from public offers.
- Responders who accepted offers decide how much to return back to the Proposers.
- Payoffs are calculated as follows:
    - If an offer is made and accepted:
        - Proposer payoff = 10 (initial points) - transfer + 4 × return
        - Responder payoff = 10 (initial points) + transfer – return
    - Otherwise Proposer and Responder receive 10 points.
- The game continues on to another round with an 80% probability. Otherwise the game ends, and a new game begins.

After 10 minutes, with the first new game to start, we will add a new part to the game to be explained then. Your decisions in the games before that do not affect your payoffs in this additional part.

After the tutorial we will ask you to solve some exercises to help you get accustomed to the experiment. Following this phase we will begin the experiment, which will last between 1 to 1.5 hours

Further instructions

We will ask you some additional questions while you're playing the game.

Questions for Proposers

The first question

The first question asks about how well you expect to do in future games. Specifically:

How much do you expect to earn on average *per round* in the *next game*?

Note that this is **not** a question about the **current** game. It is your best estimate on what your payoff per round will be in the **next** game. You can make use of the payoff tables distributed in order to come up with an estimate; please turn to these now.

Points earned will be calculated according to the formula below:

Points = $9 - [(3/80) \times (\text{estimate} - \text{actual})]^2$

This formula awards you for the accuracy of your estimate. For example,

- Assume your estimate was 30, and the actual value turned out to be 30 as well. You will earn 9 points.
- Assume your estimate was -30, and the actual value turned out to be 50. You will earn 0 points.

The second question

The second question asks how much return you expect if your offer is accepted. If you have made an offer, you will be asked to guess:

How much do you expect the Responder who accepts your offer to return to you, assuming that your offer is accepted?

Unlike the previous question, this asks you specifically how you expect a Responder who accepts your **current** offer to behave.

Points will be calculated according to the formula:

Points = $9 - [(3/10) \times (\text{estimate} - \text{actual})]^2$

As in the previous question, this formula awards you for the accuracy of your estimate.

Question for Responders

As you can see above, we ask Proposers every round:

How much do you expect to earn on average *per round* in the *next game*?

They earn extra points according to how accurate their expectations are. We want to ask **you** the following:

How do you think the Proposer you played with last round answered this question?

Note that this is NOT a question about the **current** game. Rather, it is your best estimate of her beliefs about the **next** game. You can make use of the payoff tables distributed in order to come up with an estimate.

You will earn points depending on how close your estimate is to the actual answer of the Proposer you played with in the previous round. Your points will then be calculated according to the formula:

Points = $9 - [(3/80) \times (\text{your estimate} - \text{actual answer})]^2$

++++++++++++++++++

You will be asked these questions only if they apply to you. Otherwise you will obtain a payoff of 9 points for sure.

For all of these questions the best payoff you can receive is 9 points, while the worst you can receive is 0 points. You will receive 0 points if you do not answer the questions. Since your prediction is made before actual values are determined, the best thing you can do to maximize your expected payoff from prediction is to simply state your true beliefs.

You will be paid only for estimates from the **final** round of each game. The probability that the current round is the final one, is 20% no matter which round you are currently in. Points earned from belief questions will be calculated, reported and paid to you at the end of the experiment. We will stop asking you about beliefs after 70 minutes.