Flexibility and Reputation in Repeated Prisoner's Dilemma Games

Dorothée Honhon
Jindal School of Management, University of Texas at Dallas, Dorothee.Honhon@utdallas.edu,
http://www.hyndman-honhon.com/Main/Dorothee_Honhon.html

Kyle Hyndman
Jindal School of Management, University of Texas at Dallas, KyleB.Hyndman@utdallas.edu,
http://www.hyndman-honhon.com/Main/Kyle_Hyndman.html

We study how three matching institutions, differing in how relationships are dissolved, affect cooperation in a repeated prisoner's dilemma and how cooperation rates are affected by the presence of a reputation mechanism. While cooperation is theoretically sustainable under all institutions, we show experimentally that cooperation rates are lowest under random matching, highest under fixed matching and intermediate in a flexible matching institution, where subjects have the option to dissolve relationships. Our results also suggest important interactions between the matching institution and reputation mechanism. Under both the random matching and flexible matching institution, both subjective (based on subjects' ratings) and objective (based on subjects' actions) reputation mechanisms lead to substantial increases in cooperative behavior. However, under fixed matching, only the subjective reputation mechanism leads to higher cooperation. We argue that these differences are due to different reputation mechanisms being more forgiving of early deviations from cooperation under certain matching institutions, which gives subjects the ability to learn the value of cooperation rather than getting stuck with a bad reputation and, consequently, uncooperative relationships.

Key words: Long-term relationships; prisoner's dilemma; reputation
OR/MS subject classification: C72, C91, D83
History: **July 22, 2019**

1. Introduction

Many important business and personal relationships involve repeated interaction over a short or long-term basis. Toyota's underlying philosophy of *keiretsu* – building networks of suppliers in long-term relationships that work closely to identify cost-savings and product improvements – is one of the pillars of the Toyota Production System and has been linked to its success in the auto market (Liker and Choi 2004). For the garment industry, Uzzi (1996) demonstrates empirically that firms which receive only short-term contracts were substantially more likely to go out of business than firms with long-term contracts. In general, while many relationships have no fixed end date, the parties engaged typically have the option to terminate their partnership should things sour.

In practice, 50% or more partnerships/joint ventures end in failure (Coopers and Lybrand 1986, Bamford et al. 2004, Anderson and Jap 2005, KPMG 2009) and nearly half of marriages end in divorce (Aughinbaugh et al. 2013).

Although there are many reasons why relationships break down, in this paper, we focus on the players' incentives to behave opportunistically. Examples of opportunistic behavior in business partnerships include actions by Microsoft, which, prior to their acquisition of Nokia, limited the compatibility of its then newest mobile operating system to only the newest Nokia phones, thus hurting Nokia (Grundberg and Stoll (2012)). As another example, Neuville (1997) describes opportunistic behavior by an automobile parts supplier which was eventually discovered to have been cutting corners with respect to quality.

A growing trend in business transactions is the increased reliance on reputation mechanisms as a way to provide information about the parties' potential for a successful relationship. A number of leading companies such as Boeing, Home Depot and Wal-Mart use "supplier score cards" to measure the performance and effectiveness of their vendors (see e.g., Doolen et al. 2006, Shokoohyar 2018, and the references therin). Similarly, most e-commerce platforms include some form of a rating system: Amazon Marketplace reports the percentage of positive feedback received by a seller; Etsy, TripAdvisor, Yelp and Airbnb all publish a 5-point star rating system calculated as an average from past user scores and travel website Orbitz now offers a score out of 10 on flights as an average measure of objective characteristics such as flight duration and the amenities offered on board. Ebay recently announced a shift from subjective ratings to more objective measures.[1] These reputation scores constitute an important tool for deciding with whom to engage in a relationship and for how long There is also growing awareness of the need to tailor the reputation mechanism to the underlying market platform: Bolton et al. (2013) and Nosko and Tadelis (2015) study how the reputation system interacts with the eBay trading platform and Fradkin et al. (2018) consider the Airbnb vacation rental platform. For example, in one-shot interactions, such as a purchase on eBay or an apartment rental on Airbnb, the buyer may be most interested in learning whether in the past, packages were delivered on time or whether the amenities offered at the property matched the promised list. In contrast, for a recurrent service such as lawn/landscaping, housecleaning, or even many subscription meal delivery services, subjective measures may be more relevant, such as courtesy and trustworthiness. To be effective, a reputation mechanism, should convey the relevant information given the scope of the interaction. What works for one-time purchases, may not work for (potentially) long-term services.

In this paper, we seek to understand the implications that the ability to terminate ongoing relationships has on cooperative behavior when the parties involved can behave opportunistically,

[1] See, for example, `http://pages.ebay.com/sellerinformation/news/fallupdate2015/index.html`.

as in the classical prisoner's dilemma. We also study the impact of reputation mechanisms on cooperation. To do this, we conduct laboratory experiments where subjects play the prisoner's dilemma game repeatedly under three different matching institutions which vary based on how relationships are dissolved, as well as under three forms of reputation mechanisms. In terms of matching institution, one extreme is the *Indefinitely Binding Agreements* (IBA) institution in which subjects are matched in relationships of indefinite duration. The other extreme is the *random matching* (RM) institution in which new relationships are formed every period. In between these two extremes we consider the *Temporarily Binding Agreements* (TBA) institution, where, at the end of each period subjects have the option to dissolve the relationship – and be subsequently rematched from within the pool of subjects whose relationships dissolved – or to maintain the relationship.

We study the above mentioned matching institutions with three reputation mechanisms: (i) no reputation; (ii) an *objective* scoring system; and (iii) a *subjective* scoring system. With the objective mechanism, subjects are informed of the average frequency of cooperation of their match in all previous rounds, while with the subjective mechanism, subjects observe the average satisfaction rating of their match across all previous matchings, which is a score on a 5-point scale given by subjects to their matches upon the dissolution of their matching.

In theory, cooperation is sustainable as an equilibrium in all matching institutions provided players are sufficiently patient. For the IBA case, Friedman (1971) showed that cooperation can be sustained with simple trigger strategies. For the RM case, Kandori (1992) showed that cooperation can be sustained via a community enforcement mechanism. Specifically, a defection by one player sets off a contagion towards universal defection by the population. Anticipating this, players are deterred from defecting. Using a similar, though simpler logic, we show that cooperation is theoretically sustainable – under the same conditions as indefinite repetition – when players have the option to dissolve relationships in the TBA institution. Subjects are deterred from defection by the same trigger strategies as under indefinite repetition and the belief that subjects in the rematching pool (should the relationship be dissolved) are uncooperative.

Despite cooperation being theoretically possible in all matching institutions, our first experimental result is that cooperation generally increases going from the RM to TBA to IBA institution. Compared to RM, we argue that the TBA institution facilitates cooperation by allowing a *sorting* process between subjects to occur. In particular, the relationships that start off cooperative are maintained, while other relationships (where players defect) dissolve. Contrary to what the theory predicts, we find that the rematching pool is not completely uncooperative, so that new cooperative relationships can be formed following a dissolution while progressively leaving uncooperative subjects amongst themselves to engage in short-lived partnerships with mutual defections.

Although, in theory, reputation mechanisms play no role in sustaining the most cooperative equilibrium, we find that, for the RM and TBA matching institutions, both objective and subjective reputation mechanisms lead to significantly higher levels of cooperation than the baseline case without a reputation mechanism – by at least 70% and as much as 225%. In contrast, only a subjective reputation mechanism is effective at increasing cooperation rates in the IBA institution; cooperation is actually *lower* with an objective reputation. We argue that this is because the subjective reputation mechanism is *less* forgiving of defections than the objective reputation mechanism under TBA but *more* forgiving under IBA and RM. With a less forgiving reputation mechanism, subjects who are initially uncooperative may get "stuck" with a bad reputation, which leads to future uncooperative relationships. In the absence of a reputation mechanism or with a more forgiving reputation mechanism, these same subjects can learn to cooperate. In other words, a more forgiving mechanism allows subjects to learn the value of cooperation and the importance of a good reputation, thereby facilitating cooperation over the long-run. This suggests that there are interesting interactions between the matching institution and the reputation mechanism so that it is important to choose the right reputation mechanism for a given matching institution.

2. Related Experimental Literature

There is a growing experimental literature which studies cooperation in the indefinitely repeated prisoner's dilemma. Roth and Murnighan (1978) is the first such study; they introduced the methodology of a random termination in order to mimic an infinitely repeated game. Roth and Murnighan (1978) along with Murnighan and Roth (1983) provided early evidence that cooperation can be sustained under indefinite repetition provided that it is a subgame perfect equilibrium. More recent papers, such as Dal Bó (2005) and Dal Bó and Fréchette (2011) provided more nuanced results on when cooperation can be expected to emerge. Dal Bó and Fréchette (2011) show that cooperation is further enhanced if, in addition to being subgame perfect, it is also risk-dominant. Duffy and Ochs (2009) are interested in how the matching protocol affects cooperation and focus on comparing fixed matching with random matching environments. We take these as two end points on a continuum and also consider an intermediate case where players have the option to dissolve relationships (TBA).

A growing number of papers are interested in either the endogenous termination or formation of groups, in both prisoner's dilemma and public goods game settings. Early research by Orbell and Dawes (1993) showed that voluntary participation into the prisoner's dilemma led to increased cooperation in one-shot settings as the value of the outside outside option went up, while Hauk (2003) showed that this result extends to the finitely repeated prisoner's dilemma. More recently, Nosenzo and Tufano (2017) also showed that voluntary participation increases cooperation. Furthermore, their research was able to clarify that the threat of non-participation was the main driver,

rather than a belief that cooperators are more likely to participate, thereby making voluntary participation a kind of sorting mechanism.

More closely related are papers by Zhang et al. (2016), Lei et al. (2018) and Lee (2018) who all study prisoner's dilemma games where players have the option to terminate relationships. Zhang et al. (2016) is similar to us in that they show that cooperation increases as it becomes more difficult to dissolve relationships; however, their experimental implementation differs from ours in terms of what subjects believe about the rematching pool.[2] Lei et al. (2018) focuses only on temporarily binding agreements using a similar design as the former paper and are focused on identifying strategies of players. They show that rather than dissolving a relationship following a defection, many subjects maintain the relationship but behave similarly to how they would behave against a new match. Finally, Lee (2018) shows that costless separation (as in our TBA treatment) does not harm cooperation, though the baseline level of cooperation in her IBA setting is quite low. She also shows that introducing a cost of dissolving a relationship can actually increase cooperation, but payoffs net of separation costs do not necessarily increase. None of these papers study how reputation mechanisms affect cooperation.

Similar to us and the above-mentioned papers, Wilson and Wu (2017) consider an indefinitely repeated partnership game where subjects have the option to terminate their relationship (as in our TBA institution), but unlike in our TBA institution, when a subject terminates a relationship, the two partners receive an exogenous termination payoff in each subsequent period until the end of the repeated game. They show that the relationship between efficiency and the termination payoff is non-monotonic. Initially, cooperation rates increase with the value of the termination payoff, which increases efficiency. However, for sufficiently high (but inefficient) termination payoffs, subjects more frequently terminate relationships and cooperation rates do not improve, leading to a decline in efficiency. In our TBA institution, subjects are rematched upon the dissolution of a matching, so that termination payoffs can be seen as endogenous, and we show a strong relationship between these endogenous termination payoffs and a subject's reputation.

Outside of the prisoner's dilemma, there are also several related public goods game studies. In the same spirit as Orbell and Dawes (1993), Keser and Montmarquette (2011) show that contributions to a public good are higher when subjects can voluntarily choose to be paid based on team effort. Similarly, My and Chalvignac (2010) show that contributions increase with the outside option. Gaudeul et al. (2015) study flexibility in a public goods game in which players can exit and contribute to a private good or maintain the relationship and contribute to a public good. In

[2] Specifically, within the same supergame, the rematching pool contains subjects who endogenously terminated their matching and also some subjects whose matching was *exogenously* terminated.

their paper, the decision to exit is reversible. They show that subjects exit excessively, leading to inefficiency; moreover, they find that introducing barriers to exiting improves welfare.

As we will show, in the cooperative equilibrium, subjects should never defect and should, therefore, never dissolve relationships. However, off the equilibrium path, the ability to dissolve relationships could act as a sorting device.[3] In a indefinitely repeated gift-exchange game, Bernard et al. (2018) also show that flexibility promotes sorting such that, eventually, more cooperative types are more likely to be matched with each other and less cooperative types are more likely to be matched with each other.

We also contribute to the literature on reputation in repeated games. Duffy and Ochs (2009), Camera and Casari (2009) and Kamei (2017) study reputation by varying the information available about subjects' past actions in repeated prisoner's dilemma games and show that they can enhance cooperation. Stahl (2013) also considers reputation in a repeated prisoner's dilemma, using a color coded system, which changes between green and purple as a function of past behavior, as a signal of reputation. We complement these papers by examining reputation mechanisms in an environment where relationships may be prematurely dissolved by the players, i.e., TBA. Reputation has also been studied in other contexts. Duffy et al. (2013) examine reputation in trust games, while Bolton et al. (2004, 2005, 2013) study at reputation in online trading markets. In all of these cases, reputation mechanisms are found to be efficiency-enhancing. With the exception of Bolton et al. (2013), all of the reputation mechanisms are based on objective information about subjects' past behavior. In contrast, our study considers both an objective and a subjective reputation mechanism.

Finally, Kamei and Putterman (2017) study (objective) reputation mechanisms in which subjects can form new groups in each period of a finitely repeated public goods game based on the observable reputation information. Subjects play multiple finitely repeated games and the reputation information is wiped clean at the start of each repeated game. They show that subjects learn the value of reputation so that contributions in early periods increase across instances of the finitely repeated game. Although we focus on reputation mechanisms which are long-lasting, we document that different reputation mechanisms are more forgiving of early uncooperative behavior depending on the matching institution, which, like Kamei and Putterman (2017), allows subjects to learn to cooperate and to learn the value of reputation.[4]

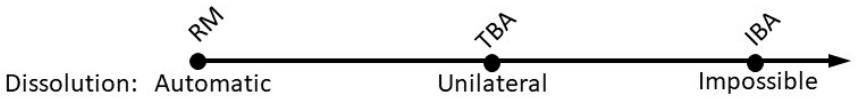
[3] Indeed, in Section 7.1, we show using a simple model that in a world with conditional cooperators and subjects who always defect, the flexibility to dissolve promotes sorting of subjects by player type.

[4] In the e-companion, we also briefly report the results of an experiment where subjects' reputations were periodically wiped clean. However, unlike Kamei and Putterman (2017), cooperation rates did not appear to increase.

3. Equilibrium Analysis

As our unit of theoretical analysis, we consider a *supergame* to be the indefinite repetition of a prisoner's dilemma stage game (e.g., as in Figure 2) where, at the end of each *period*, the supergame continues with probability $\delta \in (0,1)$ and terminates with probability $1-\delta$.[5]

Figure 1 A Taxonomy of the Matching Institutions Governing Relationship Dissolution



Our first research question concerns how the matching institution, and in particular, the ability to dissolve a relationship, affects pairs of subjects' ability to sustain cooperation in these indefinitely repeated games. We distinguish between three matching institutions as depicted in Figure 1. At the left-most extreme, players are engaged in a series of one-shot interactions and dissolutions are automatic at the end of every period. We refer to this as *Random Matching* (RM). On the right-most extreme, under *Indefinitely Binding Agreements* (IBA), players are indefinitely bound together for the entire duration of the supergame so that dissolutions are impossible. In the middle, under *Temporarily Binding Agreements* (TBA), subjects have the option to dissolve their relationship at the end of each period and one player is enough to trigger a dissolution.[6] We assume that upon dissolution of a relationship, players are randomly rematched from within the group of players whose relationships dissolved in that period. We will refer to this group of players as the *rematching pool*. Our goal in this section is to provide, for each institution, conditions such that mutual cooperation is sustainable in equilibrium. The analysis makes use of (grim) trigger strategies to show the existence of cooperative equilibria because of their tractability. There are many other strategies that can also sustain cooperation in equilibrium and, also, many equilibria which do not achieve full cooperation.

Indefinitely Binding Agreements (IBA). Let π_{ab} denote the payoff to a player when she chooses action $a \in \{C, D\}$ and her opponent chooses action $b \in \{C, D\}$ in the prisoner's dilemma. In order to sustain mutual cooperation as a subgame perfect equilibrium (using grim-trigger strategies), the expected total payoff from mutual cooperation until the end of the supergame, which is equal to $\frac{\pi_{CC}}{1-\delta}$, should be greater or equal than the expected total payoff from defecting in the first period, followed by mutual defection thereafter, which is equal to $\pi_{DC} + \frac{\pi_{DD}\delta}{1-\delta}$. Hence, we require:

$$\frac{\pi_{CC}}{1-\delta} \geq \pi_{DC} + \frac{\pi_{DD}\delta}{1-\delta} \Leftrightarrow \delta \geq \frac{\pi_{DC} - \pi_{CC}}{\pi_{DC} - \pi_{DD}}.$$

[5] This is equivalent to the more familiar notion of an *infinitely* repeated game, where δ represents the discount factor.

[6] In the e-companion, we also briefly discuss a fourth level, which we call TBA-M, that lies between the TBA and IBA institutions. Like the TBA institution, relationships can be dissolved; however, to do so requires mutual consent.

For the payoff matrix in Figure 2, this condition is equivalent to $\delta \geq 0.4$. In our experiments, we set the continuation probability at 90%. Therefore, mutual cooperation is a not only a subgame perfect Nash equilibrium, but also it is the risk-dominant equilibrium (see, e.g., Dal Bó and Fréchette (2011)).

Random Matching (RM). Kandori (1992) was the first to provide conditions such that cooperation may be sustained even under random matching via the community enforcement strategy. The crucial determinants are the payoffs, continuation probability and group size. Following the method of Duffy and Ochs (2009), one can show that for the parameters that we implement in our experiment, cooperation is an equilibrium of our game.

Temporarily Binding Agreements (TBA). We now show that cooperation can be sustained under the same conditions on δ as in the IBA treatment, whether or not mutual consent is required to dissolve relationships. Matters are slightly more complicated because each period can be broken into two stages: the prisoner's dilemma stage, in which players simultaneously choose an action and the dissolution stage in which players simultaneously choose whether to dissolve their relationship or not. Let $a_{i,t}$ denote the action of player i in time t and let $a_{j(i),t}$ denote the action of player i's match at time t. Let $m_{i,t}$ denote the number of times player i's relationship has dissolved from period 1 to period $t-1$, where we define $m_{i,1}=0$. For our equilibrium construction, we make use of the following partial histories: at the PD stage, for $t=1$, $h^{PD}_{i,1}=(\varnothing; m_{i,1})$ and for $t>1$, $h^{PD}_{i,t}=((a_{i,s},a_{j(i),s})^{t-1}_{s=1}; m_{i,t})$; and at the dissolution stage, for $t=1$, $h^{D}_{i,1}=((a_{i,1},a_{j(i),1}); m_{i,1})$ and for $t>1$, $h^{D}_{i,t}=((a_{i,s},a_{j(i),s})^{t}_{s=1}; m_{i,t})$. That is, we keep track of the complete history of actions and the number of times a player's relationships have dissolved up to that period.

Consider then the following strategy:

1. For the initial history, $h^{PD}_{i,1}=(\varnothing;0)$, choose `Cooperate` (C).
2. In the PD stage for $t>1$, `Cooperate` (C) if $h^{PD}_{i,t}=((C,C)^{t-1}_{s=1};0)$, otherwise `Defect` (D).[7]
3. For *any* history, $h^{D}_{i,t}$ choose `Maintain` in the dissolution stage.

One can see that this induces the same equilibrium path as in the IBA institution and any deviation from the equilibrium path leads to mutual defection thereafter. Hence, this strategy will be a subgame perfect equilibrium under the same conditions as IBA. Moreover, it does not matter whether unilateral or mutual consent is required to dissolve a relationship. This is because, under the given strategy, players never exercise their option to dissolve a relationship.

Of course, there are many other possible equilibrium strategies that can sustain cooperation. For example, we can modify part (3) of the above strategy as follows:

[7] In our experiment, the (random) end of the supergame is the same for all players. Therefore, subjects know that any time they are rematched within the same supergame, they are rematched with another subject whose relationship was dissolved.

3′. In the dissolution stage, choose `Maintain` if $(a_{i,t}, a_{j(i),t}) = (C, C)$, otherwise `Dissolve`.

It also induces the same equilibrium path but involves subjects exercising their option to dissolve following a defection by either player in the relationship. The key intuition behind these equilibria is that any defections are punished by defection forever thereafter and the rematching pool is uncooperative. Anticipating this, players choose to cooperate in every period and relationships never dissolve.

Reputation Mechanisms. We are also interested on the effects of reputation mechanisms on cooperation for the various matching institutions. Of course, since cooperation is already sustainable without a reputation mechanism, it remains an equilibrium when one is present. Therefore, in the fully cooperative equilibrium, along the equilibrium path, reputation mechanisms are not beneficial. However, there is reason to believe that, off the equilibrium path, reputation mechanisms do affect cooperation and that they may have different effects depending on the matching institution. We defer a discussion of this to Section 5.

4. Experimental Design

In all of our experiments, each session was divided into an a priori unknown number of supergames. At the beginning of each supergame, subjects were matched in groups of two and played the prisoner's dilemma game depicted in Figure 2 for an indefinite number of periods.[8] In what follows we refer to a pair of subjects as a *matching* and to a player's opponent as his or her *match*. We followed standard procedures, first used by Roth and Murnighan (1978) to implement an indefinitely repeated game in a laboratory setting. At the end of each period, the experimental software draws a random number between 0 and 1: if the number is below 0.9, the supergame would continue; otherwise, it would end and all matchings would terminate. The software would then check whether 75 or more periods since the beginning of the experiment had been played. If so, the experiment would end; if not, another supergame would begin.

Figure 2 The prisoner's Dilemma Stage Game

	Cooperate (C)	Defect (D)
Cooperate (C)	$40, 40$	$12, 50$
Defect (D)	$50, 12$	$25, 25$

We used a 3×3 experimental design. In one dimension, we varied the matching institution by varying how relationships are dissolved in each supergame: automatically each period (RM), upon the request of one or both subjects (TBA) or never (IBA). In the other dimension, we varied

[8] To avoid any framing effects, we used neutral language in the experiment. In particular, the stage-game actions were labeled "Action A" and "Action B" rather than "Cooperate" and "Defect".

the type of reputation mechanism that was available: none, objective reputation and subjective reputation. Table 1 provides the number of sessions, the number of participating subjects, the number of periods as well as average subject earnings for each treatment.

In the IBA matching institution, subjects remained paired with the same match for the entire duration of each supergame. In contrast, for the TBA matching institution, each period of a supergame was divided into two stages. In the first stage, subjects simultaneously chose an action, while in the second stage, subjects observed the action of their match and their respective payoffs, then were asked if they wished to remain paired with the same subject or be rematched. As noted, it was sufficient for one member of the matching to request to be rematched to trigger the dissolution of the matching. Assuming that the supergame continued for another period, then all subjects whose matchings dissolved would enter the rematching pool and new matchings would be formed from subjects in this pool.[9] Note that subjects were not given the identity (or any kind of identifying information) of their match.

In the random matching (RM) institution all subjects in the session were randomly rematched after *every* period and not just at the end of every supergame. This is consistent with how Duffy and Ochs (2009) implemented their random matching treatments.

To investigate the role of reputation mechanisms in facilitating cooperation, we consider two mechanisms: Objective (Obj) and Subjective (Sub) reputation. In the objective reputation mechanism, subjects received the following information at the start of each new period for both themselves and their match: (i) the average frequencies of cooperation and defection, overall (i.e, from the beginning of the experiment), (ii) the average frequencies of cooperation and defection in the first period of a new matching,[10] and (iii) the total number of matchings so far (which, in RM, is also the total number of periods since the start of the experiment).

In the subjective reputation mechanism, subjects rated their level of satisfaction with their match on a 5-point Likert scale – where 1 indicated *completely dissatisfied* and 5 indicated *completely satisfied* – which we call *satisfaction score*, each time their matching dissolved (i.e., at the end of each period in RM, at the end of each supergame in IBA and when requested as well as at the end of each supergame in TBA).[11] We show in Section 8 that subjects tend to rate partners who

[9] To the extent that it was possible, the software tried to ensure that subjects would be matched with a different partner from one period to the next. However, if only one group dissolved in a given period, by design the same two subjects would be rematched again in the next period. In this event, both subjects were told that they had been randomly rematched among the pool of subjects whose relationship had broken up in the previous period and no mention was made that it was actually the same match. This happened for less than 1%, 2.8% and 4.7% of the matching-periods combinations respectively when no reputation mechanism was present, with subjective and objective reputation.

[10] Because (i) and (ii) are identical in the RM matching institution, we only showed one metric.

[11] In our data analysis, the satisfaction scores were re-scaled to $[0, 1]$ to make it comparable to the other reputation treatments.

are more cooperative more favorably and partners who are less cooperative less favorably, but that there are interesting nuances that lead to differences across matching institutions. Note that both objective and subjective scores were calculated since the start of the experiment, i.e., the numbers were not reset at the start of a supergame.

Table 1 Summary of Treatments

Matching	Reputation	Sessions	# of Subjects	# of Periods	Average Earnings	Session Dates
RM	None	3	20, 18, 20	79, 75, 88	\$16.03	Sep 2016
RM	Objective	3	20, 18, 18	85, 77, 84	\$16.64	Nov 2018
RM	Subjective	3	14, 20, 18	75, 84, 75	\$16.69	Nov 2018
TBA	None	3	16, 16, 14	81, 83, 90	\$18.76	Apr 2014
TBA	Objective	3	20, 20, 18	75, 76, 88	\$21.02	Nov 2014
TBA	Subjective	3	20, 18, 20	97, 77, 94	\$21.92	Nov 2014
IBA	None	3	16, 16, 16	88, 98, 87	\$21.81	Apr 2014
IBA	Objective	3	12, 16, 14	110, 76, 91	\$21.56	Nov 2018
IBA	Subjective	3	14, 14, 16	81, 78, 75	\$19.85	Nov 2018

For each treatment, we ran three sessions at the experimental laboratory of a public university in the United States.[12] In total, 462 subjects participated, of which approximately 2/3 were male. All subjects were students – undergraduate and masters level – currently registered at the university. Because of the random termination rule, sessions varied in length, with the shortest session lasting for 75 periods and the longest session continuing for 110 periods. No session lasted for more than 90 minutes. The subjects' average earnings (including a \$4 participation fee) for each treatment are given in Table 1. Note that there is variation in earnings both due to the treatment condition and the different number of periods each session had. The experiment was programmed in z-Tree (Fischbacher 2007). Sample instructions for the TBA treatment can be found in EC.3.

5. Hypothesis Formulation

Given the previous equilibrium analysis, we see that cooperation is, at least theoretically, possible under all of our treatment conditions. Therefore, we state our **null hypothesis** as:

H_0: *Cooperation rates are identical across all treatment conditions.*

Of course, while full cooperation is *theoretically possible* in all treatments, there are multiple equilibria, not all of which are fully cooperative. Furthermore, as several authors have shown (cf. Duffy and Ochs (2009), Dal Bó and Fréchette (2011)), just because cooperation is an equilibrium does not mean that it will be achieved in an experiment. We first consider what we believe is the

[12] As can be seen in Table 1, there is a substantial time difference between when the reputation treatments for the TBA matching institution and the reputation treatments for the IBA and RM matching institutions were conducted. Although we have no reason to suspect that behavior vis-à-vis reputation mechanisms would change in the intervening years, we cannot rule it out. We refer the reader to Fisman et al. (2015) for a discussion of how distributional preferences may have changed as a result of the Great Recession.

most plausible **alternative hypothesis** in the absence of a reputation mechanism. We then discuss how reputation mechanisms may affect our alternative hypothesis. In what follows, let C_M denote the average frequency of cooperation in matching institution M, where $M \in \{RM, TBA, IBA\}$.

The existing experimental literature on whether community enforcement can sustain cooperation under random matching is mixed. While Camera and Casari (2009) show that it is possible to sustain cooperation, Duffy and Ochs (2009) come to the opposite conclusion. There are three main differences between these two papers. The former paper considers cohorts of size 4, has a continuation probability of $\delta = 0.95$ and the payoff matrix is such that trigger strategies can support cooperation (in IBA) if $\delta \geq 0.25$. In contrast, the latter paper has larger groups (6 or 14), a continuation probability of $\delta = 0.9$ and can support cooperation under IBA if $\delta \geq 0.5$. Given that our set-up is closer to Duffy and Ochs (2009) (groups of size 14 or 16, $\delta = 0.9$ and cooperation can be sustained for $\delta \geq 0.4$), we expect results closer to theirs – namely, that cooperation will be lower in RM than IBA. Hence, $C_{RM} < C_{IBA}$.

While the strategies that support cooperation under temporarily binding agreements are less cognitively demanding than Kandori (1992), they still require that subjects anticipate that a defection will lead to subsequent defection and, moreover, that the rematching pool consists of players who will defect. To the extent that players believe that some players in the rematching pool may cooperate, it increases the temptation to defect on their current match because they can dissolve the relationship (to avoid punishment in the form of future defections) and potentially start a new and cooperative matching. Thus $C_{TBA} < C_{IBA}$. At the same time, since a relationship has the potential to be of indefinite duration, it is likely that at least some subjects will behave cooperatively. Hence, $C_{RM} < C_{TBA}$. Therefore we state our hypothesis as follows:

H_a: (i) *In the absence of a reputation mechanism, cooperation rates are lowest under RM and highest under IBA, that is,* $C_{RM} < C_{TBA} < C_{IBA}$.

Consider now what happens to cooperation when a reputation mechanism is present. We first note that there is good reason to expect that the subjective mechanism will be effective. Subjects have the incentive to provide accurate feedback, in the sense of a high (low) score for a match that cooperated (defected), because of the likelihood that they will be rematched with the same match in some future period or supergame. Indeed, Bolton et al. (2013) show that subjective feedback systems can be efficiency-enhancing in buyer-seller markets.[13]

In random matching environments, several studies (e.g., Bolton et al. 2004, 2005, Duffy and Ochs 2009, Camera and Casari 2009, Duffy et al. 2013) show that providing objective information

[13] Bolton et al. (2013) compared a "traditional" subjective feedback system with variants designed to address some of the inherent problems in the traditional setting such as retaliatory feedback. They did not compare such systems with one in which feedback is not possible at all. However, in similar settings, Bolton et al. (2004, 2005) showed that introducing an objective feedback systems is efficiency-enhancing.

about subjects' past behavior increases trust, trustworthiness and cooperation. These studies show a range of effect sizes, from quite small (e.g., Duffy and Ochs (2009)) to quite large (e.g., Bolton et al. (2004)). Compared to these studies, the subjects in our experiments have even more of an incentive to cooperate because defection in one supergame will negatively affect one's reputation at start of the next supergame, which could reduce the likelihood of subsequent cooperation, while, for example, Bolton et al. (2004) consisted of a single 30-period long supergame. Because of this, we believe that the presence of a reputation mechanism should enhance cooperation in all three institutions we consider.

Hence, we have the following *alternative hypothesis* regarding reputation mechanisms:

H_a: (ii) *Adding a reputation mechanism will increase cooperation in all matching institutions.*

While we believe that cooperation will be more frequent with a reputation mechanism, there are also reasons to be skeptical of their efficacy. For example, in the IBA institution, Duffy and Ochs (2009) show that cooperation rates are typically increasing across supergames because subjects tend to initially under-estimate the benefits of cooperation. Therefore, it is possible that many subjects could start the second supergame with a relatively bad reputation, despite them having learned in the first supergame that cooperation is preferable. As a result, these subjects may become victims of their past behavior and unable to form new, cooperative, relationships, leading to lower cooperation rates than without a reputation mechanism. A similar logic could also apply to the TBA institution, especially if the rematching pool is completely uncooperative. This suggests that exactly how effective a reputation mechanism is in promoting cooperation may depend on the matching institution and on the subject's behavior off the equilibrium path.

At the same time, the type of reputation mechanism (i.e., objective vs subjective) may impact cooperation rates differently for a given matching institution. As noted, we believe that the subjective reputation will be effective at promoting cooperation, but, because different subjects may have different rating criteria, the subjective measure is likely to be a noisy signal of cooperativeness. Although feedback in our experiment is given simultaneously, it is also possible that subjects may be spiteful. For example, consider the outcomes (C, C) and (D, C) to the prisoner dilemma's game, where the first letter is the action of the subject and the second that of her match. In both cases, the subject's match cooperated; yet, the subject may choose to rate her match differently because her payoff was different. This creates further noise, which could also weaken the informativeness of the mechanism, and imply that an objective measure is a more accurate measure of past behavior.

Another important dimension to consider is the frequency with which subjects's reputation scores are updated relative to the length of a relationship. Under the objective reputation mechanism, the reputation scores are updated at the end of each period under all three institutions. Under the subjective mechanism, the updates occur only when new scores are given, which is upon the

dissolution of a relationship. Therefore, these changes occur systematically at the end of each period in the RM treatment; occasionally upon the dissolution of a matching or at the end of a supergame in TBA and only at the end of a supergame in IBA. This potentially has an impact on how relatively more forgiving/punishing a reputation mechanism can be in a given matching institution. Consider the IBA institution and suppose that a relationship was largely uncooperative. Here, the objective measure would punish both subjects because of the long history of defection. However, in the subjective treatment, subjects may be somewhat *forgiving*, meaning that they would not assign too low of a score to their match at the end of the supergame. In contrast, consider a relationship with history $\{(C,C),\ldots,(C,C),(D,C)\}$ in the TBA institution which ended after the single defection. In this case, it is quite possible that the subject who cooperated could give a low subjective satisfaction score to punish his match for defecting in the last period. However, the objective reputation would be more forgiving because of the long history of prior cooperation.

The above discussion suggests that there are reasons for and against both mechanisms, and that it might be dependent on the matching institution. Therefore, we state the following research questions:

Q: *(i) Are reputation mechanisms equally effective across matching institutions? (ii) Is one reputation mechanism (objective or subjective) universally better than the other, or does it depend on the matching institution?*

6. General Results on Cooperation Rates

In this section we test our null hypothesis of equal cooperation across all treatments against our conjectured alternative hypotheses that (i) without reputation, cooperation rates increase going from RM to TBA to IBA and (ii) that introducing a reputation mechanism leads to higher cooperation.

In Table 2 we report the average frequency of cooperation, i.e., picking action C in the prisoner's dilemma phase for each treatment for each of our treatments, organized according to our 3×3 experimental design where we vary the matching institution and the reputation mechanism. In different panels, we differentiate between different cuts of the data, looking separately at cooperation in (i) the first period of the first supergame, (ii) the first period of all supergames, (iii) all periods of the first supergame and (iv) all periods of all supergames. This allows us to comment on the pattern of cooperation rate both across and within supergames.

The table reports two sets of tests. First, to examine whether cooperation increases as we go from RM to TBA to IBA, we report the $p-$values of the non-parametric trends test (based on session averages) separately for each reputation mechanism.

Our first observation is that, although there is no clear relationship between cooperation and the matching institution in the first period of the first supergame (top-left panel), when we consider

Table 2 Average Cooperation Rates (in %)

First SG; First Period				**All SGs; First Period**			
	Reputation				Reputation		
Matching	None	Objective	Subjective	Matching	None	Objective	Subjective
RM	43.10	55.36	61.54***	RM	18.61	21.69	33.99***
TBA	65.22	79.31	75.86	TBA	36.18	77.86***	59.04*
IBA	64.58	57.14	65.91	IBA	63.86	62.95	75.48**
$p-$value (NP Trend)	0.155	0.881	0.707	$p-$value (NP Trend)	0.037	0.100	0.025

First SG; All Periods				**All SGs; All Periods**			
	Reputation				Reputation		
Matching	None	Objective	Subjective	Matching	None	Objective	Subjective
RM	12.01	27.78***	33.47***	RM	7.46	14.40**	24.29***
TBA	60.92	81.72	84.03	TBA	39.11	83.01***	66.56**
IBA	42.71	58.37	49.56	IBA	62.64	57.90	71.75***
$p-$value (NP Trend)	0.037	0.297	0.101	$p-$value (NP Trend)	0.011	0.180	0.025

Note 1: *, ** and *** denote significance at the 10%, 5% and 1% levels respectively of the estimated marginal effect of introducing either an objective or subjective reputation mechanism in the given matching institution. For each component of the table we estimated a model where the decision to cooperate was the dependent variable and we had a full set of interactions between matching institution and reputation mechanism as explanatory variables. Except in the case of the first period of the first supergame (upper-left portion of the table), we estimated random-effects models. In all cases, the standard errors are robust to clustering at the session level.
Note 2: The Non-Parametric Trend test takes the session average as the unit of independent observation.

all periods of all supergames, stronger patterns emerge (bottom-right panel). In particular, when no reputation is available or when a subjective reputation mechanism is available, cooperation increases going from RM to TBA to IBA; that is, $C_{RM} < C_{TBA} < C_{IBA}$. In contrast, with the objective reputation mechanism, cooperation rates are highest in the TBA institution.

Next, for each matching institution, we test whether cooperation is different when comparing the case of no reputation mechanism to each of the possible reputation mechanisms. In a similar vein, we see that differences take time to emerge but when we consider all periods of all supergames (bottom-right panel), we find that for all matching institutions, cooperation is significantly higher when a subjective reputation mechanism is present than when no reputation is available. In contrast, while the objective reputation mechanism significantly increases cooperation for the RM and TBA institutions, there is no significant difference between cooperation rates in the IBA institution with or without an objective reputation mechanism.

We summarize these findings in the following:

Result 1 *Consistent with our alternative hypothesis ($H_a(i)$), for both no reputation and subjective reputation mechanism, over the long-run we find that $C_{RM} < C_{TBA} < C_{IBA}$. However, contrary to $H_a(i)$, under an objective reputation mechanism, cooperation rates are highest under the TBA institution.*

Furthermore, consistent with $H_a(ii)$, across all matching institutions, a subjective reputation mechanism yields significantly higher cooperation over the long-run compared to no reputation mechanism. In partial contrast to $H_a(ii)$, there does not appear to be any benefit to adding an objective reputation mechanism in the IBA institution.

In Section 8 we return to a discussion of the reputation mechanisms and argue that the reason for these results is that the subjective mechanism is more forgiving of defections in the RM and IBA institutions, but less so in the TBA institution.

7. Temporary Binding Agreements: Cooperation, Flexibility and Reputation

As discussed in Section 3, the key equilibrium logic behind the TBA institution is that subjects should cooperate at the start of each supergame and continue to cooperate so long as they have always experienced mutual cooperation. Off the equilibrium path, once a subject's relationship breaks up, she should defect in every subsequent period. In this section we focus on the TBA institution and look at cooperation decisions at the start of supergames and upon rematching, as well as the decision to dissolve relationships. We are also interested in how the presence of a reputation mechanism affects these decisions.

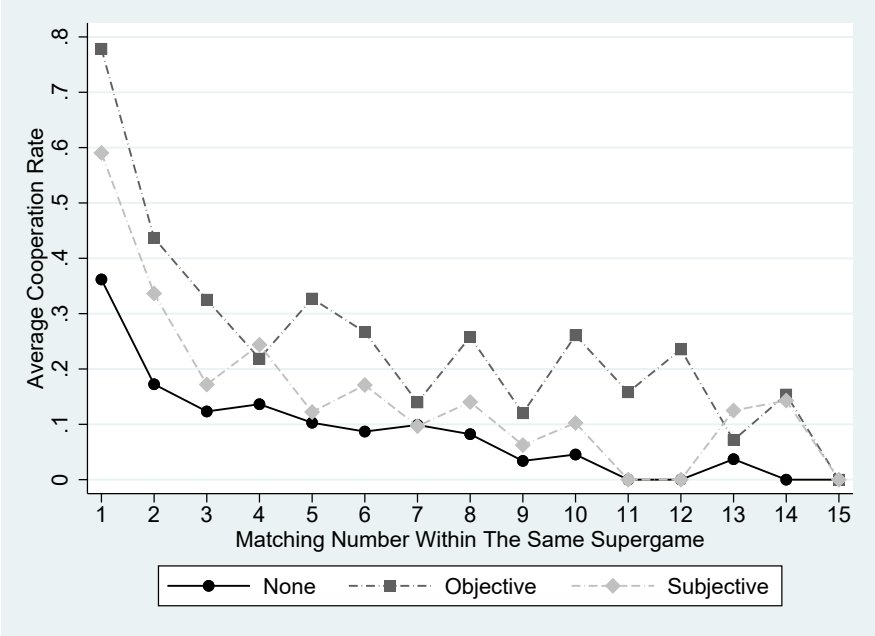
7.1. Cooperation and the Rematching Pool

Figure 3 plots the average cooperation rate in the TBA treatments in the first period of a new matching as a function of the matching number (within each supergame). As can be seen, contrary to the theoretical prediction, across all reputation treatments, far from all subjects cooperate in the first matching within a supergame and, in the subsequent matchings, the cooperation rates are still mostly positive, though declining. In the absence of a reputation mechanism, the cooperation rate in the first period of a new matching appears to hit a value very close to zero only after about 11 matchings.

There are two factors that may explain this downward trend. First, since relationships that start off with mutual cooperation are likely to be maintained, the subjects who end up in the rematching pool in later periods are likely to be the ones who defect. Second, because of the declining potential for cooperation of the subject pool, even a subject who is open to cooperating may choose to defect if she is sufficiently pessimistic about the likelihood of meeting another cooperator.

To illustrate this, we propose two simple models. In the first model, we assume that the subject pool consists of two types of players: *defectors* and *(unconditional) cooperators*. Defectors are subjects who defect and dissolve relationships in every period. Cooperators are players who cooperate in every period and who maintain their relationship if the outcome was (C, C), otherwise they

Figure 3 The Evolution of Cooperation in the First Period of Matchings Within the Same Supergame (TBA Institution)



Note: We only include data up to the first 15 matchings that a subject had in a supergame. This represents 96% of the relevant data.

dissolve. Let $\hat{p} \in [0,1]$ denote the proportion of cooperators in the subject pool ($1-\hat{p}$ being the proportion of defectors) and p_t denote the expected proportion of cooperators in the *rematching pool* at the start of period t of a supergame. For $t>1$, the rematching pool is composed of the subjects whose matching broke up in period $t-1$.

On average, a fraction $1-\hat{p}^2$ of subjects will see their first matching in a supergame dissolve (because at least one of the player was a defector) and, therefore, will find themselves in the rematching pool at the start of period 2. The remaining expected proportion $(\hat{p})^2$ of subjects are cooperators who were matched with another cooperator and will therefore cooperate and maintain their relationship until the end of the supergame. As a result, the expected proportion of cooperators in the rematching pool at the start of period 2 is $p_2 = \frac{\hat{p}-\hat{p}^2}{1-\hat{p}^2} = \frac{\hat{p}}{1+\hat{p}}$. By extending this argument to period t, we have:

$$p_t = \begin{cases} \hat{p}, & \text{if } t=1, \\ \frac{p_{t-1}}{1+p_{t-1}} = \frac{\hat{p}}{1+(t-1)\hat{p}} & \text{if } t>1. \end{cases} \quad (1)$$

This simple model captures the first factor explaining the decline in cooperation rates on Figure 3: as t increases, the expected proportion of subjects choosing to cooperate at the start of a new matching, which is equal to the proportion of cooperators in the rematching pool, is decreasing.

We now propose a second model to incorporate the second factor; we now assume that cooperators are *conditional cooperators*; that is, they only cooperate in the first period of a matching if it is in their best interest to do so, in the sense that their expected discounted profit is higher if they

choose action 'C' than action 'D'. These players continue to cooperate with a given match and maintain their relationship as long as the outcome is (C,C) and dissolve the relationship otherwise. We further assume that the conditional cooperators are aware of their relative proportion in the subject pool, i.e., $\hat{p}$. The remaining $1-\hat{p}$ players are still defectors as described above.

Let $V(C,p_t)$ and $V(D,p_t)$ denote the expected discounted total profit that a conditional cooperator receives from cooperating and defecting, respectively, if the current proportion of conditional cooperators in the rematching pool is p_t, assuming that they will start defecting in period $t+1$ if their match defects in period t. We have:

$$V(C,p_t) = \frac{\pi_{CC}p_t}{1-\delta} + (1-p_t)\left(\pi_{CD} + \frac{\pi_{DD}\delta}{1-\delta}\right) \tag{2}$$

$$V(D,p_t) = p_t\left(\pi_{DC} + \frac{\pi_{DD}\delta}{1-\delta}\right) + \frac{\pi_{DD}(1-p_t)}{1-\delta} \tag{3}$$

Conditional cooperators choose to cooperate if and only if $V(C,p_t) \geq V(D,p_t)$ which is equivalent to $p_t \geq p^* = \frac{(\pi_{DD}-\pi_{CD})(1-\delta)}{\pi_{CC}+\pi_{DD}-\pi_{CD}-\pi_{DC}+\delta(\pi_{CD}+\pi_{DC}-2\pi_{DD})}$.

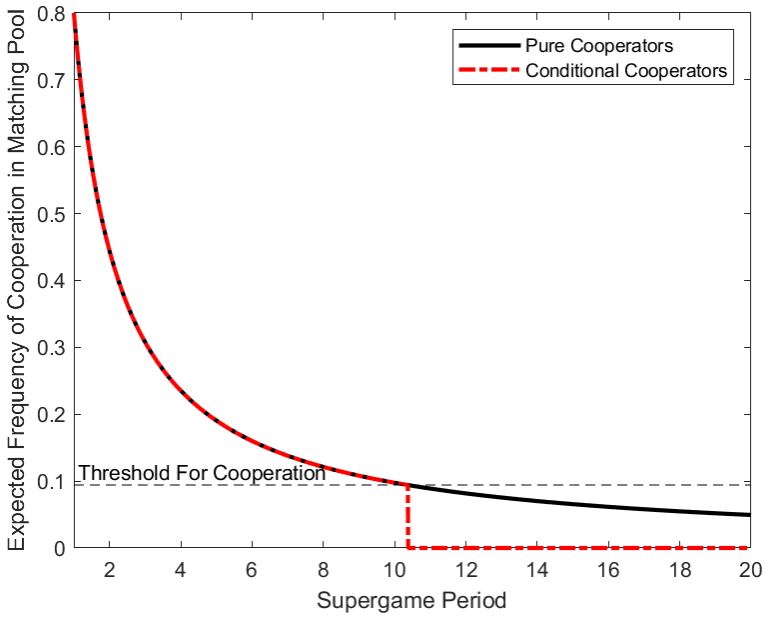
Using (1), we get that conditional cooperators will cooperate in period t if $t \leq t^* = 1 + \frac{\hat{p}-p^*}{p^*\hat{p}}$ and will defect otherwise. As as result the cooperation rate in the first period of a matching will be equal to p_t from (1) for $t \leq t^*$ and zero for $t > t^*$.

In our game where $\pi_{CC} = 40, \pi_{DD} = 25, \pi_{CD} = 12, \pi_{DC} = 50$ and $\delta = 0.9$, we have $p^* \approx 9.42\%$, which implies that $t^* = 11.62 - 1/\hat{p}$. Figure 4 represents the evaluation of the expected proportion of cooperators for $\hat{p} = 0.8$, such that $t^* \approx 10.37$, implying that conditional cooperators should cooperate in the first period of the first 10 matchings they experience then defect afterwards.[14] If their match is also a conditional cooperator, they will form a stable relationship until the end of the supergame, otherwise, if their match is a defector, their relationship will break up and both will enter the rematching pool. We depict the expected evolution of cooperation rates at the start of new matching in Figure 4 for our models with (unconditional) cooperators and conditional cooperators. The observed patterns appear similar to the experiment data from Figure 3.

Remark 1. While a formal analysis is beyond the scope of the paper, we also provide some intuition on how a reputation mechanism would alter the behavior of conditional cooperators. Suppose that the reputation mechanism provides a binary signal (good or bad), with accuracy $\theta \in [0.5, 1]$, about a subject's true type (conditional cooperator or defector). A subject in period t, with prior belief p_t on the fraction of conditional cooperators left in the matching pool, who receives a signal that his match is a conditional cooperator updates his belief to $\tilde{p}_t = {}^{\theta p_t}/_{(\theta p_t + (1-\theta)(1-p_t))}$. Substituting $\tilde{p}_t$ into (2) and (3), one can again solve for the latest period, $\tilde{t}$, that a conditional cooperator would

[14] Looking at Figure 3, a plausible range for $\hat{p}$ is between 0.4 and 0.8, in which case t^* varies between 9.12 and 10.37. Indeed, except for very small values of $\hat{p}$, t^* does not vary much.

Figure 4 The Evolution of Cooperators in the Matching Pool



cooperate with a new match. Observe that for a completely uninformative reputation mechanism, $\theta = 0.5$, we obtain $\tilde{t} = t^*$, as derived above. However, as the signal becomes more informative, i.e., θ tends to 1, we get $\tilde{t} \to \infty$; that is, as the reputation mechanism becomes perfectly informative, a conditional cooperator is *always* willing to cooperate with a new match with a "good" reputation.

This simple framework suggests that there are potentially three main types of subjects: (1) *Unconditional cooperators*, who always cooperate in the first round of a new matching; (2) *Conditional cooperators* who cooperate in the first round of a new matching for matchings 1 to K but defect in the first round of a new match for matchings $> K$ (where K may vary across individuals of this type, and may be higher in the presence of a reputation mechanism); and (3) *Defectors*, who always defect in the first round of a new matching. Next we use our experimental data to estimate the relative proportion of subjects of each type and classify those who do not fit the definitions above as "other". Table 3 contains the frequency of observing each type of subject.

Table 3 Classification of Types By Behavior in Round 1 of Each Match

Reputation	Unconditional Coop.	Conditional Coop.	Defectors	Other
None	23.5%	3.9%	59.3%	13.3%
Objective	69.8%	3.6%	10.4%	16.2%
Subjective	52.1%	5.2%	23.7%	18.9%

Note: Unconditional Cooperators are subjects who always cooperated in the first round of each new match in the supergame. Conditional cooperators are those subjects who cooperated in the first round of each new match until some threshold, and then defected in subsequent matches. Subjects are classified as Defectors if they always chose to defect in the first round of a new match in the supergame. Other captures subjects who display any other behavior.

First, observe that unconditional cooperators, conditional cooperators and defectors make up

about 85% of the player types, and only about 15% of the time do subjects display some other type of behavior (e.g., switching from defecting to cooperating in the first round of successive matches at least once), which suggests that our simple models are a good representation for actual subject behavior. Second, we see that the presence of a reputation mechanism substantially increases the fraction of subjects who are willing to cooperate and reduces the fraction of subjects who always defect. Finally, it appears that conditional cooperators are a minority of the population of cooperative subjects, never comprising more than 5% of the overall population.[15] Note also that if we look at cooperators of any kind who experienced at least 2 matches, then the split between conditional and unconditional cooperators is much more balanced.

We summarize the results of this discussion as follows:

Result 2 *The rematching pool becomes less cooperative over the course of a supergame. This appears to be driven by two factors: (1) some cooperators leave the rematching pool and (2) some conditional cooperators eventually stop cooperating.*

By comparing the difference in reputation scores at the beginning of the first matching and last matching of a supergame, we are also able to verify another prediction of our simple model of (conditional) cooperators and defectors: we show that the flexibility of TBA promotes sorting. Because the effect is economically small, we relegate a full discussion of this to EC.1. However, we summarize the result as follows:

Result 3 *In the TBA matching institution, subjects are more likely to be matched with more similar counterparts in terms of cooperative behavior at the end of the supergame than at the beginning.*

7.2. The Decision To Dissolve a Relationship

In our equilibrium analysis of Section 3, we showed that cooperation can be sustained under a variety of assumptions on dissolution behavior once a defection has occurred – subjects could either choose to maintain relationships and punish defectors by defecting themselves or dissolve relationships and enter into a rematching pool in which players always defect. As our simple models of (conditional) cooperators and defectors in §7.1 suggest, dissolution upon experiencing a defection may be a better decision when subjects can expect to find a cooperative relationship in the rematching pool, especially in the presence of a reputation mechanism since a defection

[15] Because of censoring, it is difficult to fully distinguish between unconditional and conditional cooperators. It is likely that some truly conditional cooperators found a stable match early and so were never tempted to begin a new matching by defecting and, consequently, are labeled as unconditional cooperators according to our classification. Therefore, it may make more sense to think of the sum of conditional and unconditional cooperative subjects in the table as being "potentially cooperative" and to be somewhat agnostic about the exact mix between the two types.

directly (for objective) or indirectly (for subjective) affects one's reputation score. In this section, we document three key findings: (i) subjects are very likely to dissolve a relationship following a defection by one or more players; (ii) more cooperative subjects are more likely to dissolve following a defection; and (iii) it is generally disadvantageous to dissolve a relationship, except for subjects with an observably high (objective or subjective) reputation.

Table 4 shows the percentage of matchings that are dissolved conditional on the outcome of the prisoner's dilemma phase. Not surprisingly, when both subjects cooperate, less than 1% of matchings dissolve. When one subject defects, in all reputation conditions, approximately 45% of matchings are dissolved. Interestingly, in these cases, when the relationship dissolves, between 72 and 74% of the time, the subject who cooperated is the one to request the dissolution. This indicates that many prefer to take their chances with a new match rather than enter into a punishment phase or try to build a cooperative relationship with the current match who defected. Lastly, when both subjects defect, a strong majority of the time, the matching is dissolved at the request of one or both subjects. We also see that there is a large difference in behavior between the objective and subjective reputation mechanisms. We conjecture that the differences are due to subjects employing different strategies to try to protect their reputation. Specifically, in the objective mechanism, subjects anticipate the negative unavoidable impact that future defections with the same partner would have on their objective reputation score and prefer to break the relationship for a greater chance at cooperation with another match. On the other hand, in the subjective mechanisms, if subjects are pessimistic about the rematching pool it may be better to postpone the assignment of the subjective satisfaction scores, preferring a sequence of (D, D) with the same partner and only one negative rating to a sequence of low scores received from several successive partners.

Table 4 Frequency of Matchings That Dissolve By Prisoner's Dilemma Game Outcome

Prisoner's Dilemma Outcome	None	Objective	Subjective
Both Subjects Cooperate	0.18%	0.94%	0.58%
One Subject Cooperates, The Other Defects	48.18%	47.06%	44.16%
Both Subjects Defect	64.14%	84.92%	59.55%

In Table 5, we report the results of a regression analysis on the decision to terminate a relationship for each of the three reputation mechanisms. Since dissolution occurs less than 1% of the time following mutual cooperation, we condition on at least one defection. In the absence of a reputation mechanism ("None" column), we include the subject's frequency of cooperation as their reputation.[16]

[16] This is the same measure as in the Objective mechanism, with the difference being that it was not displayed to subjects. To the extent that we expect that subjects to be aware of their own frequency of cooperation we believe that it is reasonable to include it as an explanatory variable.

The main takeaway from this analysis is that more cooperative subjects (i.e., those with a higher reputation score) are more likely to dissolve a relationship following a defection. However, we also see that the effect is stronger in the two reputation treatments, which suggests that subjects with higher reputation believe that dissolving matchings may be advantageous to them. Another interesting finding from the Objective and Subjective columns is that subjects are less likely to dissolve the higher the reputation of their match.

Table 5 The Role of Reputation in the Decision to Terminate a Relationship (Conditional on At Least One Defection)

	None		Objective		Subjective	
(Own Action, Match's Action)						
(C, D)	0.059	(0.064)	0.079^{*}	(0.048)	0.020	(0.049)
(D, D)	0.266^{***}	(0.051)	0.349^{***}	(0.074)	0.179^{***}	(0.046)
Subject's Reputation	0.383^{***}	(0.087)	0.464^{***}	(0.084)	0.560^{***}	(0.117)
Subject's Number of Matchings	−0.002	(0.003)	0.001	(0.003)	−0.004	(0.003)
Index of Supergame	−0.001	(0.007)	−0.002	(0.016)	0.001	(0.015)
Match's Reputation			−0.041	(0.095)	-0.152^{**}	(0.066)
Match's Number of Matchings			0.005^{**}	(0.002)	0.002	(0.001)
Constant	0.103^{*}	(0.055)	−0.052	(0.077)	0.122^{*}	(0.074)
R^2	0.074		0.188		0.115	
Observations	2700		996		2010	

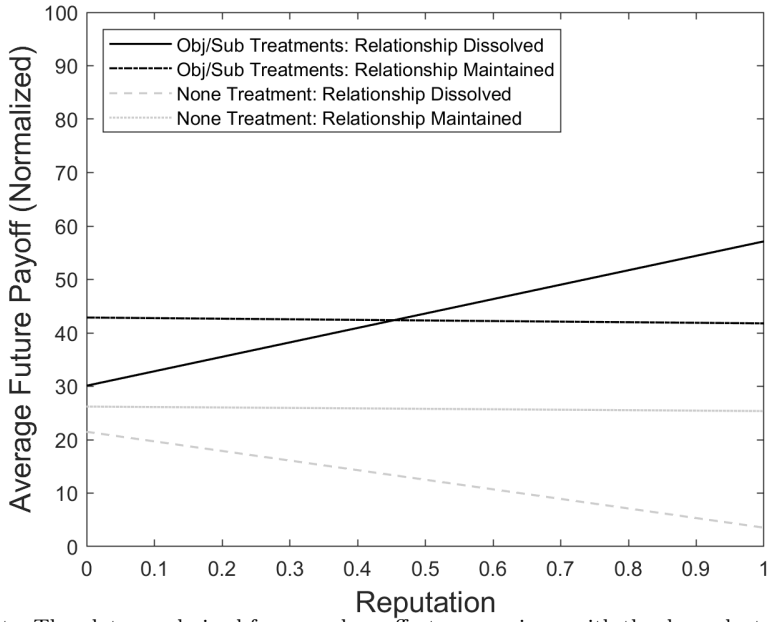
Note 1: *, ** and *** denote significance at the 10, 5 and 1% levels, respectively. The table reports linear random effects models with bootstrapped standard errors in parentheses.
Note 2: In all cases, the dependent variable is an indicator variable, which takes value 1 if the subject chose to dissolve the relationship and 0 otherwise.
Note 3: The variable "Reputation" is on a continuous scale from 0 to 1, with 1 indicating a "perfect" reputation. In "None" and "Objective" reputation is the overall frequency of cooperation. To be sure, in "None", this information was not displayed but a subject could reasonably be expected to be aware of their own average frequency of cooperation; hence, we include it as a comparison with the actual reputation treatments. In "Subjective", the reputation variable corresponds to the (normalized) average satisfaction score.

Finally, we investigate when the decision to dissolve a relationship pays off. In Figure 5 we plot the normalized predicted average payoffs until the end of the supergame, conditional on at least one defection, as a function of a subject's reputation (or frequency of cooperation absent a reputation mechanism).[17]

This figure provides three main insights. First, reinforcing previous results, a reputation mechanism is efficiency-enhancing when subjects have the flexibility to dissolve relationships. As can be seen, average future payoffs are always higher when a reputation mechanism is present. Second, in the presence of a reputation mechanism, the future payoff from dissolving a relationship is strictly increasing in one's reputation – so much so that subjects with high enough reputation (above about 0.45 on the normalized scale) actually benefit from dissolving relationships (following at least one defection). Third, absent a reputation mechanism, subjects are always better-off maintaining a

[17] The average future payoff from period t is computed as: $\bar{\pi} = (1/T-t)\sum_{i=t+1}^{T} \pi_i$, where T is the period in which the supergame ends. We then normalize this so that it is reported as the percentage of maximum gains from cooperation. That is, $\bar{\pi}^N = 100((\bar{\pi}-25)/(40-25))$.

Figure 5 Average Normalized Payoffs in Remainder of Supergame (Conditional on At Least One Defection)



Note: The plots are derived from random effects regressions, with the dependent variable being the average normalized future payoff and the explanatory variables subjects' reputation, an indicator for whether the relationship was maintained and their interaction. For the treatments without a reputation mechanism we use frequency of cooperation as a proxy. EC.1 provides additional details, including estimated marginal effects of reputation on future payoffs for each possible stage game/dissolution outcome.

relationship, regardless of their frequency of cooperation (which we take as a proxy of their reputation) and, in fact, the future payoff from dissolving a relationship is actually decreasing in one's cooperation rate.

Given the analysis of this section, we can summarize our result as:

Result 4 *Subjects are more likely to dissolve following at least one defection and this is especially true of subjects with higher reputations when such a mechanism exists. In the absence of a reputation mechanism, subjects who maintain relationships following a defection earn more, on average, than those who dissolve. This effect disappears when a reputation mechanism is available because high reputation subjects are more likely to form cooperative relationships in the rematching pool following a dissolution and, consequently, earn more.*

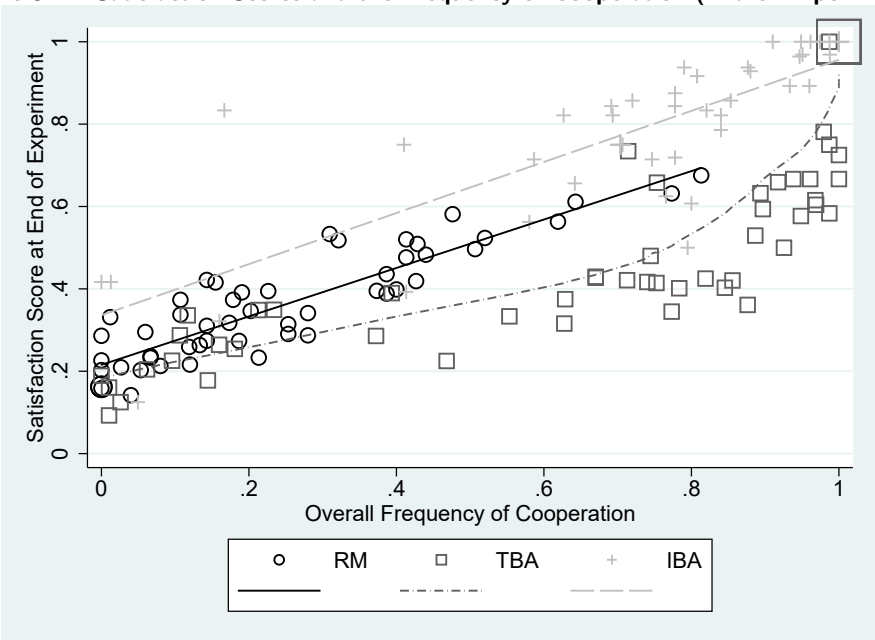
8. Subjective Reputation and Behavior

Our results in §6 show that a subjective reputation mechanism increases cooperation over the long-run in all three matching institutions. Moreover, in the RM and IBA institutions, it actually leads to higher cooperation rates than the objective reputation mechanism (see Table 2). Given this, it is interesting to understand how a subject's subjective reputation score, i.e., the average satisfaction score given by her matches on a 1 to 5 scale, correlates with cooperative behavior

and how subjects rate each other following the dissolution of a relationship under the subjective reputation mechanism under TBA. Remember that the updating frequency of the reputation scores depends on the matching institution and the reputation mechanism (objective or subjective).

For each matching institution, Figure 6 plots a subject's average satisfaction score (normalized to $[0, 1]$) in the subjective reputation treatments at the end of the experiment against his or her frequency of cooperation. As can be seen, the two variables are strongly positively associated with each other; that is, subjects who cooperated more frequently received higher satisfaction scores from their matches throughout the experiment.

Figure 6 Satisfaction Scores and the Frequency of Cooperation (End of Experiment)



Note 1: Recall that, for the analysis of results, the satisfaction scores were rescaled so that it is between 0 and 1.
Note 2: For the RM and IBA institution, the lines in the figure represent the best linear fit, while for the TBA institution – because of an apparent non-linearity, the line is derived from a non-parametric (lowess) regression. The size of each symbol represents the number of times the data point was observed.

Interestingly, in the RM and IBA institutions, the relationship appears to be linear, while the relationship is non-linear in the TBA institution. Also note that the average subjective satisfaction scores are almost always lowest in the TBA institution for a given frequency of cooperation level. This suggests that, when it comes to rating their match subjects are harshest in the TBA institution and more *forgiving* in the other two institutions. Another way to see this is to note that in RM and IBA, 86.5% and 72.7% of subjects have (normalized) reputation scores greater than or equal to their frequency of cooperation, while in TBA only 41.4% of subjects do. As we discuss further below, this may explain why the subjective reputation mechanism achieves higher cooperation rates in the RM and IBA institutions than the objective reputation mechanism.

Table 6 Average Satisfaction Scores Given in Subjective Reputation Treatments When Relationship Ends

			RM				TBA				IBA	
			Match				Match				Match	
Naturally			C	D			C	D			C	D
	Own	C	4.72	1.39	Own	C	4.76	2.23	Own	C	4.75	2.21
		D	3.75	1.99		D	2.85	2.34		D	3.14	2.62
							Match					
Other Dissolves (and Subject Maintains)							C	D				
					Own	C	5.00	1.88				
						D	3.25	2.09				
							Match					
Subject Dissolves							C	D				
					Own	C	3.78	1.30				
						D	2.60	1.91				

Note: Cells which are highlighted are those which have at least 50 observations, while cells which are not highlighted have less than 50 observations. Empty cellblocks indicate that the particular condition was not possible in a given treatment. In particular, in IBA-Sub and RM-Sub, all pairings dissolved naturally – either at the end of a supergame or every period.

It is interesting to dig deeper into how subjects rate their match upon the dissolution of a relationship and how it depends on both the actions chosen and, for the TBA matching institution, how the relationship ended. To this end, Table 6 reports the average satisfaction score for each combination of actions chosen in the prisoner's dilemma phase and for each way the relationship terminated – naturally, because of a dissolution request by the subject him/herself or because of a dissolution request by his or her match.[18] There are a several interesting findings. First, not surprisingly, subjects rate their match substantially higher when their match cooperated. Second, across matching institutions, subjects ratings are very similar, conditional on the outcome. There is only weak evidence that ratings differ across treatments following mutual defection (Kruskal-Wallis test, $p = 0.07$).

We note that subjects rate their match substantially *lower* following the outcome (D, C) than following (C, C), despite the fact that the match cooperated in both cases. Conversely, subjects are more forgiving of a defection by their opponent – i.e., give a higher rating – if they themselves defected. This type of behavior distorts the connection between reputation and cooperation; yet, our findings from §6 suggest that despite these distortions, the subjective reputation mechanism still leads to higher cooperation rates than without any mechanism, and is even better than the objective mechanism in the RM and IBA matching institutions.

We now try to explain why the subjective reputation mechanism works better than the objective mechanism under the RM and IBA institutions but the opposite is true for TBA. One possible explanation, which we alluded to in the discussion of Figure 6, is that, in RM and IBA, the

[18] If a subject chose to dissolve the relationship, she would not know whether or not her match also chose to dissolve the relationship. However, if a subject chose to maintain the relationship but the match dissolved, then she would know that it was because her match had chosen to dissolve.

Table 7 Reputation and Cooperation – First Period of Second Supergame

	Average Reputation		Frequency C		Frequency (C,C)	
Matching	Obj.	Sub.	Obj.	Sub.	Obj.	Sub.
RM	0.286	0.468	0.179	0.346	0.000	0.154
TBA	0.810	0.690	0.793	0.759	0.690	0.655
IBA	0.532	0.761	0.524	0.705	0.333	0.545

subjective mechanism is more *forgiving* than the objective measure (i.e., likely to lead to higher ratings), but the opposite is true under TBA. Consider Subjects 93 and 104 from the IBA treatment with a subjective reputation. At the end of the first supergame, which occurred after 17 periods, the subjects had cooperated one and zero times, respectively. Yet, both subjects gave each other satisfactions scores of 3 out of 5, which is substantially higher than their objective frequency of cooperation, equal to $1/17$ and 0.

Next consider Subjects 453 and 468 from the TBA treatment with a subjective reputation who were matched in the first period of the second supergame. For the first 8 periods of their relationship, both cooperated but in the 9-th period, Subject 468 defected. Because of this, Subject 453 dissolved the relationship and gave Subject 468 a satisfaction score of 1 out of 5, significantly below that subject's objective frequency of cooperation of $8/9$.

These examples provide anecdotal evidence of how forgiveness in the subjective reputation mechanism interacts with matching institution. We investigate this more formally in Table 7. The first two columns report the mean reputation scores for each matching institution under the objective and subjective reputation mechanisms in the first period of the second supergame, which is the earliest point at which subjects have a well-defined reputation in all three matching institutions.

We saw in Table 2 that, in each matching institution, subjects had approximately the same cooperation rate the first supergame whether the reputation was objective or subjective. Therefore, if the subjective ratings were a perfect translation of the objective cooperation rates, then there would be no differences in reputation scores between the objective and subjective mechanisms. The fact that we observe differences in Table 7 suggests that is not the case. In particular, we see higher scores under the subjective mechanism than the objective mechanism under RM and IBA, but in the TBA institution, the pattern is the opposite.[19]

As the other columns of Table 7 show, cooperation rates are higher and coordination on (C,C) is more frequent in the first period of the second supergame for the more forgiving reputation mechanism, and the differences are particularly large for the RM and IBA matching institutions.[20]

[19] Although not shown in Table 7, it is also true that subjects' reputations are more homogenous (i.e., less variable) under the objective reputation in TBA, while in the RM and IBA institutions, they are more homogenous under the subjective reputation mechanism.

[20] Regression results (not included) suggest that the more forgiving reputation mechanism (Objective for TBA; Subjective for IBA and RM) leads to significantly higher reputation, cooperation and coordination on (C,C), even controlling for observed cooperation rates in the first supergame.

Finally, there appears to be a difference in how subjects rate their match depending on whether they are in a "hot" or "cold" state. Specifically, define a hot (resp. cold) state as one in which the subject's match defected (resp. did not defect) in the same period as the relationship terminated. Using data from the IBA institution, where subjects cannot influence the dissolution of a relationship, we observe that subjects give significantly lower satisfaction scores in hot states than in cold states – even controlling for the overall level of cooperation. In the TBA treatment, where subjects can control the dissolution of relationships, and where defections lead to such dissolutions, it is much more likely that subjects rate their match while in a "hot" state. This provides further support for our claim that the subjective mechanism is relatively unforgiving in the TBA institution. Indeed, we summarize our findings as follows:

Result 5 *A subjective reputation mechanism is more forgiving of defections in the RM and IBA institutions, while an objective reputation mechanism is more forgiving in the TBA institutions. With more forgiving mechanisms, subjects begin matchings with higher, less variable reputations and are, consequently, more likely to cooperate and experience mutual cooperation.*

9. Concluding Remarks

In this paper, we experimentally study the interaction between matching institution (in particular, the flexibility to dissolve relationships) and reputation on cooperative behavior in a repeated prisoner's dilemma game. While cooperation is theoretically sustainable under all matching institutions regardless of whether or not a reputation mechanism is present, our results suggest that both the matching institution and presence or absence of a reputation mechanism have substantial effects on cooperation.

Our first result is that, absent a reputation mechanism, cooperation rates increase going from RM to TBA to IBA (i.e., $C_{RM} < C_{TBA} < C_{IBA}$). Compared to the one-shot random matchings, the TBA institution, where subjects have the flexibility to (unilaterally) dissolve the relationship, promotes cooperation in two ways. First, because the relationship can be, in principle, indefinite, some relationships begin cooperative, are maintained and remain cooperative. Second, in contrast to theory, some subjects whose initial matching dissolved are able to form a new, long-lasting and cooperative relationships. Thus, our second result is that the flexibility to dissolve promotes cooperation through a sorting process whereby cooperative relationships are maintained, uncooperative relationships are dissolved and some new cooperative relationships are formed.

Our third main result is that reputation mechanisms generally promote increased cooperation, with large effects observed in both the RM and TBA institution and smaller effects (if any) in the IBA institution. In the TBA institution, we show that subjects who have a higher reputation

have a higher continuation value from terminating relationships following a defection and that such subjects are, in fact, more likely to dissolve following a defection. In contrast, we show that absent a reputation mechanism, subjects would generally be better off maintaining their relationship and try to build a cooperative relationship, rather than taking their chances with the rematching pool.

Our last result highlights the importance of tailoring the reputation mechanism to the matching institution. Since subjects are likely to initially under-estimate the value of cooperating and/or the importance of a good reputation, there may be early deviations from cooperation. Our results suggest that the subjective reputation mechanism is more forgiving of deviations in the RM and IBA institutions, which allows the (relatively) low levels of cooperation that we see in the RM institution to be maintained longer and allows subjects to learn to cooperate in the IBA institution. In contrast, in the TBA institution, the objective reputation mechanism is more forgiving of opportunistic defections because it accounts for the long history of prior cooperation while the subjective mechanism may not.

With the increase in online sales and virtual interaction it is likely that reputation mechanisms are going to become even more prevalent in the future. Some companies such as Airbnb and TripAdvisor currently combine both objective and subjective ways to rate partners in a transaction so a direction for future research would be to study the behavior of subjects when both types of reputation mechanisms co-exist. Another potential avenue for further investigation would be to study cooperation when giving a rating is optional — as is most often the case in practice. Additionally, Business owners often complain about the impact of ill-intended negative reviews on review platforms like Yelp given the very large sales impact of a single bad review (Cabral and Hortaçsu 2010). Therefore, it is important that people with good reputations do not have their reputation tarnished when interacting with someone with bad intentions. It would be interesting to see if a mechanism that weights new additions to one's reputation by the reputation of the person they interacted with could be developed and whether it would be effective. Finally, in our implementation of the TBA institution subjects are always re-matched upon the dissolution of a relationship. In practice this guarantee of a re-match may not always hold and it would be interesting to study how this would impact cooperation with or without a reputation mechanism.

Acknowledgments

We would like to thank the Department Editor (Yan Chen), an anonymous associate editor and three anonymous referees for valuable comments that greatly improved the paper. Conversations with Gary Bolton, Matthew Embrey, Emin Karagozoglu and Stephen Leider are also gratefully acknowledged as is the assistance of Owen Ma and Vineetha Sivadasan for running the experiments. We gratefully acknowledge the Jindal School of Management (UT Dallas) for financial support. Kyle Hyndman also acknowledges the Institute for Research in Experimental Economics (IFREE, Small Grants Program) for additional financial support.

References

Anderson, E., S.D. Jap. 2005. The dark side of close relationships. *MIT Sloan Management Review* **46**(2) 75–82.

Aughinbaugh, Alison, Omar Robles, Hugette Sun. 2013. Marriage and divorce: Patterns by gender, race, and educational attainment. *Monthly Labor Review,* Bureau of Labor Statistics.

Bamford, James, David Ernst, David G. Fubini. 2004. Launching a world-class joint venture. *Harvard Business Review* February.

Bernard, Mark, Jack Fanning, Sevgi Yuksel. 2018. Finding cooperators: Sorting through repeated interaction. *Journal of Economic Behavior & Organization* **147** 76–94.

Bolton, Gary E., Ben Greiner, Axel Ockenfels. 2013. Engineering trust: Reciprocity in the production of reputation information. *Management Science* **59**(2) 265–285.

Bolton, Gary E., Elena Katok, Axel Ockenfels. 2004. How effective are electronic reputation mechanisms? an experimental investigation. *Management Science* **50**(11) 1587–1602.

Bolton, Gary E., Elena Katok, Axel Ockenfels. 2005. Cooperation among strangers with limited information about reputation. *Journal of Public Economics* **89**(5) 1457–1468.

Cabral, Luís, Ali Hortaçsu. 2010. The dynamics of seller reputation: Theory and evidence from ebay. *Journal of Industrial Economics* **58**(54-78).

Camera, Gabriele, Marco Casari. 2009. Cooperation among strangers under the shadow of the future. *American Economic Review* **99**(3) 979–1005.

Coopers, Lybrand. 1986. *Collaborative ventures: An emerging phenomenon in information technology*. Coopers and Lybrand, New York.

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

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

Doolen, Toni, MAJ Mike Traxler, Ken McBride. 2006. Scorecards for supplier performance improvement: Case application in a lean manufacturing organization. *Engineering Management Journal* **18**(2) 26–34.

Duffy, John, Jack Ochs. 2009. Cooperative behavior and the frequency of social interaction. *Games and Economic Behavior* **66**(2) 785–812.

Duffy, John, Huan Xie, Yong-Ju Lee. 2013. Social norms, information, and trust among strangers: Theory and evidence. *Economic Theory* **52**(2) 669–708.

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

Fisman, Raymond, Pamela Jakiela, Shachar Kariv. 2015. How did distributional preferences change during the great recession? *Journal of Public Economics* **128** 84–95.

Fradkin, Andrey, Elena Grewal, David Holtz. 2018. The determinants of online review informativeness: Evidence from field experiments on airbnb. Working Paper.

Friedman, James W. 1971. A non-cooperative equilibrium for supergames. *Review of Economic Studies* **38**(1) 1–12.

Gaudeul, Alexia, Paolo Crosetto, Gerhard Reiner. 2015. Of the stability of partnerships when individuals have outside options, or why allowing exit is inefficient. Jena Economic Research Papers #2015–001.

Grundberg, Sven, John D. Stoll. 2012. Nokia sales face more hurdles. The Wall Street Journal, `http://www.wsj.com/articles/SB10001424052702304898704577480340752531300`.

Hauk, Esther. 2003. Multipl prisoner's dilemma games with(out) an outside option: An experimental study. *Theory and Decision* **54**(3) 207–229.

Kamei, Kenju. 2017. Endogenous reputation formation under the shadow of the future. *Journal of Economic Behavior & Organization* **142** 189–204.

Kamei, Kenju, Louis Putterman. 2017. Play it again: Partner choice, reputation building and learning from finitely repeated dilemma games. *Economic Journal* **127** 1069–1095.

Kandori, Michihiro. 1992. Social norms and community enforcement. *Review of Economic Studies* **59**(1) 63–80.

Keser, Claudia, Claude Montmarquette. 2011. Voluntary versus enforced team effort. *Games* **2** 277–301.

KPMG. 2009. *Joint Ventures: A tool for growth during an economic downturn*. KPMG International.

Lee, Natalie. 2018. An experiment: Voluntary separation in indefinitely repeated prisoners dilemma game. Working Paper.

Lei, Vivian, Filip Vesley, Chun-Lei Yang. 2018. Voluntary separation as a disciplinary device for long-term cooperation: Reconciling theory with evidence. Working Paper.

Liker, Jeffrey K., Thomas Y. Choi. 2004. Building deep supplier relationships. *Harvard Business Review* December, 104–113.

Murnighan, J. Keith, Alvin E. Roth. 1983. Expecting continued play in prisoner's dilemma games: A test of several models. *Journal of Conflict Resolution* **27**(2) 279–300.

My, Kene Boun, Benoît Chalvignac. 2010. Voluntary participation and cooperation in a collective-good game. *Journal of Economic Psychology* **31**(4) 705–718.

Neuville, J. 1997. La stratégie de la confiance: Le partenariat industrial observé depuis le fournisseur. *Sociologie du Travail* **20**(3) 297–319.

Nosenzo, Daniele, Fabio Tufano. 2017. The effect of voluntary participation on cooperation. *Journal of Economic Behavior & Organization* **142** 307–319.

Nosko, Chris, Steven Tadelis. 2015. The limits of reputation in platform markets: An empirical analysis and field experiment. Working Paper.

Orbell, John M., Robyn M. Dawes. 1993. Social welfare, cooperators' advantage and the option of not playing the game. *American Sociological Review* **58**(6) 787–800.

Roth, Alvin E., J. Keith Murnighan. 1978. Equilibrium behavior and repeated play of the prisoner's dilemma. *Journal of Mathematical Psychology* **17**(2) 189–198.

Shokoohyar, Sina. 2018. Supplier performance management: A behavioral study. Ph.D. thesis, The University of Texas at Dallas.

Stahl, Dale O. 2013. An experimental test of the efficacy of simple reputation mechanisms to solve social dilemmas. *Journal of Economic Behavior & Organization* **94**(October) 116–124.

Uzzi, B. 1996. The sources and consequence of embeddedness for the economic performance of organizations: the network effect. *American Sociological Review* **61**(4) 675–698.

Wilson, Alistair J., Hong Wu. 2017. At-will relationships: How an option to walk away affects cooperation and efficiency. *Games and Economic Behavior* **102** 487–507.

Zhang, Bo-Yu, Song-Jia Fan, Cong Li, Xiu-Deng Zheng, Jian-Zhang Bao, Ross Cressman, Yi Tao. 2016. Opting out against defection leads to stable coexistence with cooperation. *Scientific Reports* **6**(35902). Https://doi.org/10.1038/srep35902.

EC.1. Supplemental Results Not Included In The Main Text

EC.1.1. Flexibility as a Sorting Mechanism

The dynamics of cooperation above suggest that the flexibility to dissolve relationships in TBA should act as a sorting device, eventually leading to the matching of (conditional) cooperators with each other in (eventually) stable relationships and a separate pool of defectors. To investigate this, we compare the differences in subjects' reputation scores with their first and last match of a supergame. Specifically, for each (subject i, supergame τ) pair, let R_i^τ be subject i's reputation score at the **beginning** of supergame τ.[21] Similarly, let $R_{j_1(i)}^\tau$ be the reputation of subject i's first match at the **beginning** of supergame τ. Finally, let $R_{j_T(i)}^\tau$ be the reputation at the **beginning** of supergame τ for subject i's last match in supergame τ. Hence, $|R_i^\tau - R_{j_1(i)}^\tau|$ is the difference in reputation scores between subject i and her first match and $|R_i^\tau - R_{j_T(i)}^\tau|$ is the difference in reputation scores between subject i and her last match in supergame τ.[22] Using these metrics, we see in Table EC.1 that, under TBA, the absolute difference in reputation scores is significantly lower at the end versus the beginning of a supergame. This indicates that subjects generally end supergames matched with someone with a more similar reputation score than they began the supergame with, which is evidence that sorting occurs. For comparison purposes, under RM, there is no statistical difference as matchings are randomly selected in each period and under IBA, the numbers are exactly the same as subjects have only one match per supergame. We also note that the results are qualitatively the same if we break down further by reputation condition. We summarize the results of this section as follows:

Result 3 *In the TBA matching institution, subjects are more likely to be matched with more similar counterparts in terms of cooperative behavior at the end of the supergame than at the beginning.*

EC.1.2. Detailed Analysis of Average Future Payoffs

In Table EC.2(a) we report the average normalized payoffs until the end of the supergame depending on whether 0, 1 or 2 subjects in a matching cooperated and whether or not the matching was maintained in the current period.[23] In the absence of a reputation mechanism, regardless of the

[21] For the case of no reputation, we take the frequency of cooperation as our measure.

[22] We hold the reputation score fixed at the initial reputation at the **beginning** of the supergame so that difference cannot be due to changes in the reputation score throughout the supergame.

[23] The average future payoff from period t is computed as: $\bar{\pi} = (1/T-t)\sum_{i=t+1}^{T} \pi_i$, where T is the period in which the supergame ends. We then normalize this so that it is reported as the percentage of maximum gains from cooperation. That is, $\bar{\pi}^N = 100((\bar{\pi}-25)/(40-25))$.

Table EC.1 Sorting and Matching Institution (Absolute Difference in Reputation Score Between Subject and Match)

	Absolute Difference in Rep Score		$p-$value
Institution	First Period	Last Period	Signed-rank Test
RM	0.161	0.161	0.859
TBA	0.301	0.260	0.018
IBA	0.291	0.291	—

Table EC.2 Average Normalized Payoffs in Remainder of Supergame Conditional on Outcome in the Previous Period

(a) (Normalized) Average Future Payoffs

	None		Objective		Subjective	
Outcome Last Period	Dissolved	Maintained	Dissolved	Maintained	Dissolved	Maintained
(D,D)	12.91%	14.40%	26.72%	39.62%	21.01%	17.27%
(C,D) or (D,C)	13.02%	42.50%	51.22%	49.04%	33.36%	35.96%
(C,C)	3.33%†	96.78%	51.45%†	97.65%	32.44%†	94.65%

† Recall from Table 4 that less than 1% of relationships dissolve following (C,C). Hence, caution is warranted in these numbers.

(b) Marginal Effects (ME) of Reputation on Average Normalized Future Payoffs Depending on Outcome

ME Computed At:					
Outcome	Matching	None		Reputation	
(D,D)	Dissolved	-10.210	(7.199)	30.117^{**}	(11.703)
(D,D)	Maintained	-1.855	(10.692)	16.173	(19.250)
(C,D) or (D,C)	Dissolved	-21.190^{*}	(11.163)	32.892^{***}	(9.423)
(C,D) or (D,C)	Maintained	-1.071	(13.213)	2.112	(13.412)
(C,C)	Dissolved	[Not estimable]		7.800	(29.640)
(C,C)	Maintained	9.620	(6.612)	4.237	(4.946)

Note 1: In panel (b), the numbers in parentheses are bootstrapped standard errors.
Note 2: In panel (b), for the "None" column, we use the same reputation score as in the objective reputation mechanism. This was obviously not observable to subjects. Hence, the fact that we see differences between the "None" and "Reputation" columns supports our claim that reputation is driving the results.
Note 3: The marginal effects are derived from linear random-effects regressions where average normalized future payoffs was the dependent variable and we had indicators for number of subjects (0, 1 or 2) in the matching who cooperated, whether the matching was maintained and the reputation score of subjects, as well as a complete set of interaction terms.

outcome in the prisoner's dilemma phase, maintaining leads to higher payoffs than dissolving. Interestingly, this is true even if one or more subjects defected and it suggests that players would be better off trying to make a relationship work than to dissolve and be paired with a subject from the rematching pool. In contrast, in the presence of a reputation mechanism, average future payoffs are essentially identical whether the relationship is maintained or dissolved when one subject defects.

These aggregate level results from panel (a) reaffirm something we saw in Figure 3 – namely, that the rematching pool is more cooperative in the presence of a reputation mechanism. However, Table EC.2(b), which shows the marginal effects of reputation on average future payoffs, indicates that these results also depend on the subject's reputation. Looking first at the reputation column we see that, when either one or two subjects in a matching defect, subjects with a higher reputation earn

significantly more from dissolving the relationship compared to maintaining it.[24] This is because, in the next matching, the likelihood of starting cooperatively (i.e., (C,C)) is significantly increasing in one's reputation and once mutual cooperation starts, it is very likely to continue.

EC.2. A Brief Analysis of Additional Treatments

EC.2.1. TBA Institution Where Mutual Consent is Required to Dissolve a Relationship

One potential problem with the TBA institution as we implemented is that a player can unilaterally dissolve a relationship. Thus, in principle, a subject who defects could escape punishment by dissolving the relationship and take his/her chances in the matching pool.[25] To see whether cooperation was affected by this feature of the matching institution, we also conducted three sessions of what we call the TBA-M institution, where relationships can only be dissolved by mutual consent. As was the case with the TBA institution, at the end of every period, subjects chose whether to maintain or dissolve the relationship. If both subjects chose to maintain, then the relationship was maintained (provided the supergame did not exogenously terminate), if both subjects chose to dissolve, then the relationship would be dissolved and, finally, if one subject chose to dissolve and the other chose to maintain, then the latter subject would be given a chance to accept or deny the dissolution request. This treatment was implemented without a reputation mechanism

We hypothesize that by making it more difficult to escape punishments, subjects will be more likely to cooperate in this institution than in the baseline TBA institution. Indeed, while the overall cooperation rate was 39.11% in the TBA institution, it was 49.42% in the TBA-M institution. Unfortunately, given only 3 sessions per treatment, we are unable to say that the difference is statistically significant.

The requirement of mutual consent has a large effect on the frequency with which relationships are dissolved. While 44.16% (resp. 59.55%) of relationships dissolve when one (resp. both) players in a relationship defect in TBA, these numbers are 14.78% and 44.86% in TBA-M. Table EC.3 shows the frequency of dissolution requests, and whether they were accepted or denied. As can be seen from the table, a substantial fraction dissolution requests are denied. This could be indicative of a desire to punish, but it is also notable that when the cooperator in a match requests a dissolution,

[24] Note also that this is not simply due to cooperative subjects having a higher reputation and, therefore, earning more by dissolving. If this were so, then we would expect to see the same results if we used the subjects' unobservable reputation score (i.e., the frequency of cooperation as in the objective mechanism). Looking at the "None" column in panel (b), it is clear that the estimated marginal effects are very different from the reputation mechanism. Indeed, a highly cooperative subject in the absence of a reputation mechanism earns significantly less in the future when a relationship is dissolved following a defection by one player.

[25] As we saw, the matching pool is generally less cooperative, so such a subject cannot entirely escape some kind of "punishment".

the subject who defected is more than twice as likely to deny the request.[26] The other interesting result is that, conditional on at least one defection, subjects are over twice as likely to cooperate (31.4%) if nobody requests to dissolve the relationship than if a request to dissolve is made but denied (13.8%). Thus subjects are willing to forgive a defection, as long as nobody tries to escape the relationship, but when one subject tries to break the relationship, the request is frequently dissolved and, in the next period, frequently punished.

Table EC.3 The Frequency of Dissolution Requests

	Outcome in Prisoner's Dilemma Phase		
Who Requests Dissolution	(C,C)	(C,D) or (D,C)	(D,D)
Neither Subject	100.00	59.35	28.21
One Subject	0.00		
Cooperator (Accepted)	n/a	9.57	
Cooperator (Denied)	n/a	19.35	
Defector (Accepted)	n/a	1.52	23.12
Defector (Denied)	n/a	6.52	26.94
Both Subjects	0.00	3.70	21.73

EC.2.2. TBA Institution With An Objective Reputation Mechanism That Resets

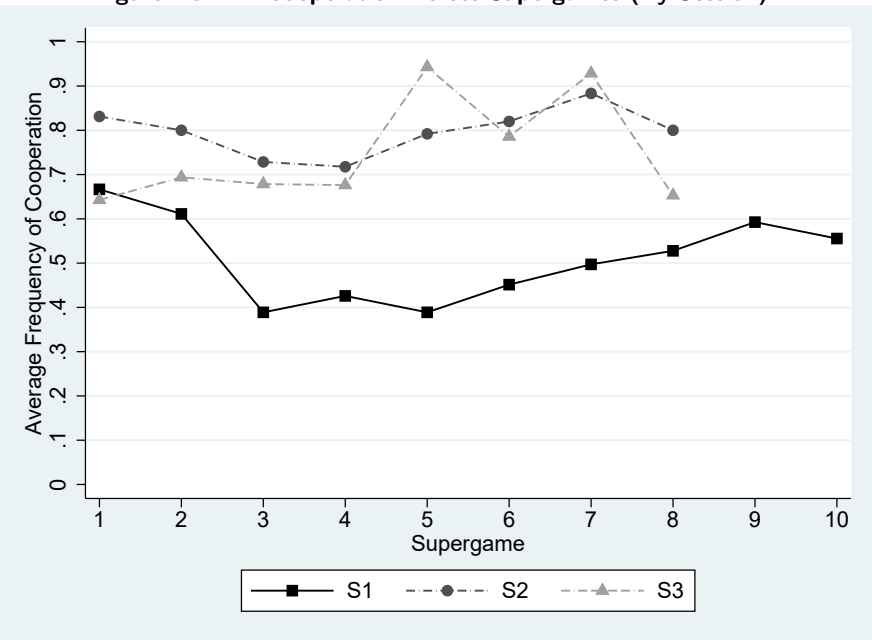
In addition to the two reputation mechanisms that we studied for our three matching institutions, we also studied behavior in the TBA institution with an objective reputation that would reset at the start of each supergame. Similar to Kamei and Putterman (2017), this allows us to examine whether subjects learn the value of maintaining a good reputation and cooperating. Specifically, because the reputations get reset at the start of each supergame a subject who defected frequently in the first supergame will not be saddled with a bad reputation in the second supergame, which should enable the subject to at least try to cooperate and build cooperative relationships.

First, observe that the average frequency of cooperation over all periods is 65.72%, which is approximately 68% higher than in the TBA institution without a reputation mechanism, and approximately equal to the overall cooperation rate in the IBA institution without a reputation mechanism but substantially less than the cooperation rate in the TBA institution with a long-lasting objective reputation. Thus, we can conclude that this mechanism facilitates cooperation in the TBA institution, though not as effectively as a long-lasting reputation.

In Figure EC.1 we plot the average frequency of cooperation for each supergame separately for each session. As can be seen, unlike Kamei and Putterman (2017), there is no clear upward trend in the frequency of cooperation across supergames. Indeed, in two sessions (S1 and S2), the

[26] Perhaps this is because the defector thinks he/she can take advantage of his/her match (because the match cooperated in the past and may do so again) or because the defector has a negative belief about the rematching pool.

Figure EC.1 Cooperation Across Supergames (By Session)



cooperation rate actually declines over the first three or four sessions before starting to increase. Therefore, at best, it takes subjects a non-trivial amount of time before they begin to realize the value of a good reputation.

EC.3. Instructions for the Temporarily Binding Agreements (Unilateral) Treatment

Thank you for coming today. If you haven't already done so, please power off all mobile devices, tablets, computers, etc and put them in your bag or on the floor at your seat. This is an experiment on the economics of decision-making. Your earnings will depend partly on your decisions and partly on the decisions of others. By following the instructions and making careful decisions you will earn varying amounts of money, which will be paid at the end of the experiment. Details of how you will make decisions and earn money are explained below.

In this experiment, you will participate in a number of decision problems (rounds). In all rounds, you will be matched with another participant in the experiment but you will not know the identity of the other participants with whom you are matched throughout the experiment. In what follows, we will refer to the person with whom you are matched as your *match* and the two of you as a *pairing*. The experiment will last for a minimum of 75 rounds. More precise details will be given below.

Decision Problem

In each round you and your match will simultaneously choose an action A or B. The payoffs for each possible combination of actions is given in the table below:

		Match's Choice	
Your Choice		A	B
A		**40**, *40*	**12**, *50*
B		**50**, *12*	**25**, *25*

The first entry in each cell (in bold) represents your payoff, while the second entry represents the payoff of the person you are matched with (in italics). As you can see, this shows the payoff associated with each choice. That is, if:

- You select **A** and your match selects **A**, you each make 40.
- You select **A** and your match selects **B**, you make 12 and your match makes 50.
- You select **B** and your match selects **A**, you make 50 and your match makes 12.
- You select **B** and your match selects **B**, you each make 25.

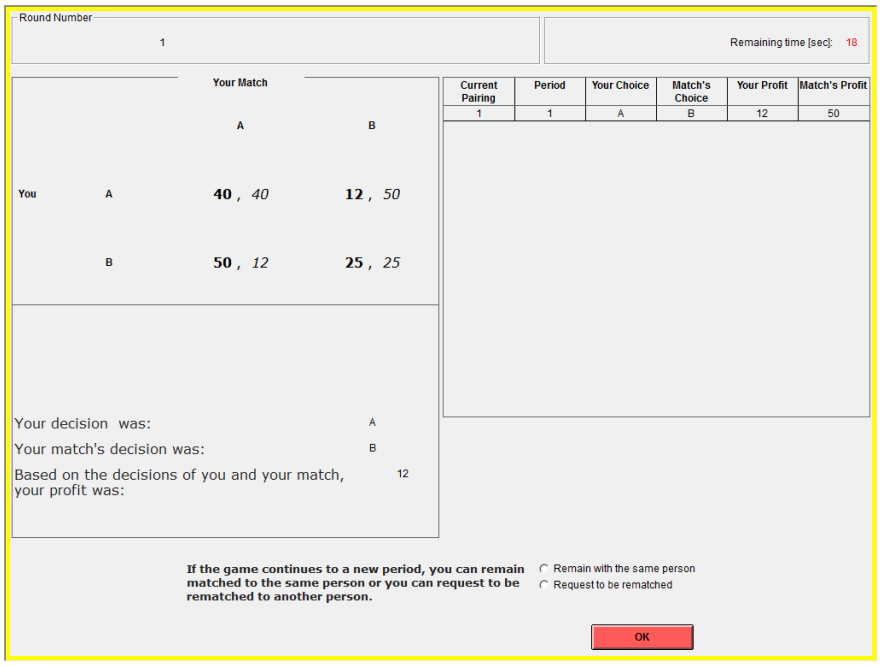
The Computer Screen

In each round, you will see the following computer screen:

On the top-left side of the screen you will see the same payoff table as depicted above. On the top-right side of the screen you can see your and your match's previous choices as well as your profits for the current pairing. That is, you will not see any information regarding choices made or outcomes from any of your previous pairings. In the example above, you see that in the first period of the third pairing, you chose **A** and your match chose **A**, which gave you a payoff of 40 and your match a payoff of 40. On the bottom-left of the screen is where you will make your decision, i.e., either action A or action B.

Payoffs

Your earnings in each round depend on your choice and on your match's choice. After both you and your match have made your choices, you will see the following screen. You see your choice, your match's choice, and your profit. In this example, you see that you chose **A**, your match chose **B** and your payoff was 12.



Pairings

At the end of every round, all participants have the option to remain matched with the same participant or to request to be rematched. If either you or your match request to be rematched, then your match for the next round will be chosen at random amongst all the participants in the experiment who either requested to be rematched or whose match requested to be rematched in the previous round.

In addition, at the end of every round, there is a 10% chance that the pairing between you and your match will **naturally** break-up. In this case, all participants will be rematched to a randomly chosen participant in the experiment and a new pairing will begin for everybody; that is, it is as if we roll a 10-sided die at the end of each round and the pairing breaks up if we roll a 1, and continues if we roll a 2, 3, . . . , 9 or 10.

Note that, absent a request for rematching, there is **always** a 90% chance that you will remain matched with the same subject in the next round. *It does not matter, for example, whether if its the first, fifth or twelfth round of your pairing; there is always a 90% chance that it will continue for one more round.*

Any time that you have been **rematched** to another participant, you will be explicitly told so; therefore, if no such announcement is made, it means you are still matched with the same participant.

End of the Experiment

The experiment will end when the first pairing **after** 75 rounds have already been played **naturally** breaks up. Note that because of the 10% chance of a natural break-up, in any round after the 75th,

you can expect the experiment to continue for approximately 10 more rounds.

At the end of the experiment, we will add all your earnings in order to determine your total points. This total will be converted to a dollar amount according to the rule:

$$\$1 = 175 \text{ points.}$$

This amount will then be added to the $4.00 participation fee to give your final payment. Payments will be made in private, in cash, after the completion of the experiment.

Rules

Please do not talk with anyone during the experiment. We ask everyone to remain silent until the end of the last decision problem.

Your participation in the experiment and any information about your earnings will be kept strictly confidential. Your receipt of payment is the only place on which your name will appear. This information will be kept confidential.

If you have any questions please ask them now. If not, we will proceed to the experiment.