A New Approach to the Analysis of Cooperation Under the Shadow of the Future: Theory and Experimental Evidence[*]

Melis Kartal[†]
Vienna University of Economics and Business
VCEE

Wieland Müller[‡]
University of Vienna & VCEE
Tilburg University, CentER & TILEC

July 31, 2018

Abstract

The theory of infinitely repeated games lacks predictive power due to equilibrium multiplicity and its insensitivity to, for example, changes in certain parameters, the timing of players' moves or communication possibilities. We propose a new approach, which mitigates the shortcomings of the theory. In particular, we study a standard infinitely repeated prisoner's dilemma game and its variants with (i) heterogeneous tastes for cooperation, and (ii) strategic risk arising from incomplete information about the opponent's taste for cooperation. We theoretically show that the variants we study reduce strategic risk and boost cooperation and coordination on cooperation relative to the standard game, unlike what a theory based on pure material self-interest of players would predict. Our theoretical results are corroborated by the results of our experiments.

JEL classification numbers: C73, C91, C92, D82, D83.

Keywords: prisoner's dilemma, cooperation, infinitely repeated games, strategic risk, game theory, experiments

[*] We would like to thank Guillaume Fréchette for helpful comments and suggestions. Furthermore, we thank Anna Osipenko for very able research assistance. The research reported in this paper is funded by the University of Vienna.

[†] Department of Economics, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria, Email: *melis.kartal@wu.ac.at*.

[‡] Department of Economics & VCEE, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, and Department of Economics, Tilburg University, CentER & TILEC, Warandelaan 2, Postbus 90153, 5000 LE Tilburg, The Netherlands. E-mail: *wieland.mueller@univie.ac.at*.

1 Introduction

The theory of repeated games has been subject to criticism as it may not provide sharp predictions due to a multiplicity of equilibria (see, e.g., Fudenberg and Maskin (1993) or Dal Bó and Fréchette (2011)). Another shortcoming of the theory is that its predictions are typically not responsive to, for example, changes in certain parameters, the timing of players' moves or communication possibilities. While the theory of infinitely repeated games has been extensively used to understand and explain cooperation in a variety of settings, this theory is firmly built on the assumption that players are exclusively motivated by their own monetary payoffs—despite the wealth of evidence showing that many people have a natural "taste for cooperation."[1] We study cooperation in the context of infinitely repeated games assuming that each player has a taste for cooperation. Our approach generates intuitive predictions that account for changes in the game, while a theory based on pure material self-interest does not. Thus, our approach mitigates shortcomings of the standard theory described above. Additionally, we present new experimental evidence supporting our predictions.

The difficulty of understanding the determinants of cooperation and accurately predicting cooperation levels in infinitely repeated prisoner's dilemma games is well-known by now in the experimental literature. It is also recognized in the literature that cooperation rates depend on the "strategic risk" of cooperation (e.g., the magnitude of the *sucker payoff* as discussed in Blonski *et al.* (2011) and the related *basin of attraction* of a defective strategy discussed in Dal Bó and Fréchette (2011)). However, a formal modeling of this strategic risk in the mental accounting of a player regarding the potential costs and benefits of cooperation is limited so far in the literature.

In our paper, cooperation and the strategic risk thereof depend not only on the stage-game payoff matrix and other related parameters implemented by the experimenter but also on certain characteristics of subjects—in particular their taste for (conditional) cooperation. We formalize this idea developing a simple model that incorporates a privately observed, heterogeneous taste for cooperation into the theory of infinitely repeated prisoner's dilemma games. This implies the following. Firstly, each player who intends to cooperate has to weigh the potential costs and benefits of cooperation estimating the chances with which the opponent has a preference for (conditional) cooperation as tastes are not observable. Secondly, players' perceptions regarding the strategic risk of cooperation are heterogeneous simply because players' tastes for cooperation are heterogeneous.

[1] Various specifications have been offered in the literature to model systematically observed deviations from material self-interest, such as preferences for (conditional) cooperation, fairness, and reciprocity. See, among others, Rabin (1993), Levine (1998), Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Charness and Rabin (2002), and Andreoni and Samuelson (2006).

In particular, even if players have similar expectations about what others plan to do, some players may have a sufficiently strong taste for (conditional) cooperation and, thus, find the strategic risk of cooperation sufficiently low, whereas those who are less cooperative, selfish or spiteful may think otherwise and prefer defection.

The strategic risk of cooperation is endogenous in our model since it depends on the (expected) fraction of players who will start by cooperating, which is determined in equilibrium. This is an important step in formalizing strategic risk as we show that this richer theory can predict behavior more successfully than the standard theory based on pure monetary self-interest of players. As a first step, we show that our comparative static predictions regarding the effect of a stage-game payoff change are consistent with recent developments in the literature. For example, a change in the sucker payoff in the standard prisoner's dilemma game changes the incidence of predicted cooperative play, which is consistent with the arguments in Dal Bó and Fréchette (2011) and Blonski *et al.* (2011), as well as the experimental evidence in Blonski *et al.* (2011) and Mermer *et al.* (2018). We then theoretically and experimentally compare cooperation and coordination rates in two variants of the infinitely repeated prisoner's dilemma game keeping the stage game payoff matrix and the discount factor identical but changing other aspects of the stage game. The first variant changes the simultaneous-move stage game to a sequential-move version, and the second variant introduces communication to the simultaneous-move stage game.

Standard theory does not differentiate between these variants, and the approaches in Dal Bó and Fréchette (2011) and Blonski *et al.* (2011) are not applicable. However, our theoretical model shows that in an infinitely repeated setting both a sequential structure and communication foster cooperation relative to the benchmark simultaneous-move game by mitigating private information problems and the risk of being the sucker. Put differently, these two variants reduce strategic risk by making it easier to recognize a cooperative opponent, and in turn this reduction in strategic risk increases the incentive to cooperate. We first compare the simultaneous game to the sequential version and predict the first-mover cooperation rate to be strictly higher than the cooperation rate in the simultaneous game. But strategic risk is the lowest for second-movers who observe first-mover cooperation, so their cooperation rate is higher than both the first-mover cooperation rate and the simultaneous game cooperation rate. Given these insights, we also predict the percentage of players who coordinate on cooperation to be strictly higher in the sequential game than in the simultaneous game. Moreover, we predict the communication game to generate an even higher rate of coordination on cooperation assuming that lying is costly to many people and thus, a sufficiently large fraction of communication is truthful (consistent with the findings of a sizable experimental literature on honesty in communication).

We test our predictions regarding the effect of sequential moves and communication relative to the benchmark simultaneous prisoner's dilemma game implementing an indefinitely repeated setting in the laboratory. The standard theory *uniquely* predicts defection in all three games. Thus, the difference between our predictions and those of standard theory are stark. Experimental results are consistent with our predictions. In particular, the communication game and the sequential game generate higher cooperation and coordination on cooperation than the benchmark simultaneous game (with the communication game generating the highest coordination and cooperation rate). As discussed above, this is because both institutions, the communication game and the sequential game, reduce strategic risk and mitigate private information and coordination problems by making it easier to recognize an opponent with a taste for cooperation.

In a nutshell, our findings suggest that theories of cooperation in infinitely repeated games may have more predictive power if they account for tastes for cooperation because such preferences can have a significant impact on behavior. Our richer setup enables us to capture an important component of strategic risk especially in those games in which cooperation is not an equilibrium or a *risk-dominant* action and allows us to make intuitive comparative static predictions regarding a change in a payoff, parameter or the format of the game. Our results also show that standard theory may have poor predictive power even if it makes a unique prediction. While standard theory uniquely predicts defection in all of our treatments, observed behavior can be far from this prediction even after players gain a large amount of experience. We also note that, to our knowledge, our study is the first one presenting a very thorough and detailed analysis of subject communication in an indefinitely repeated setting and documenting its cooperation boosting features in line with the predictions of our model.

2 Related Literature

Our paper relates to various strands of the game-theoretic and experimental literature. Firstly, our paper relates to a burgeoning experimental literature on infinitely repeated prisoner's dilemma games (see, for example, Palfrey and Rosenthal (1994), Engle-Warnick and Slonim (2004, 2006a, 2006b), Dal Bó (2005), Aoyagi and Fréchette (2009), Camera and Casari (2009), Duffy and Ochs (2009), Dal Bó and Fréchette (2011), Blonski *et al.* (2011), Fudenberg *et al.* (2012), Embrey *et* *al.* (2013), and Arechar *et al.* (2017)). Dal Bó and Fréchette (2011) and Blonski *et al.* (2011) are particularly related to our paper as they develop approaches showing the importance of strategic risk as a determinant of cooperation. Dal Bó and Fréchette (2011) analyze various infinitely repeated simultaneous prisoner's dilemma games in the laboratory. Among other things, they show that cooperation may not prevail in games in which it is a possible equilibrium action and that even if cooperation is supported in equilibrium and is risk dominant, the cooperation rate need not be high.

Our experimental findings go in a different direction showing that even if cooperation is not an equilibrium action under standard assumptions, the cooperation rate may deviate from this prediction, and in fact, it can be quite high in certain variants of the benchmark game which keeps all the parameters identical but either adds communication or introduces a sequential-move structure to the game. We show that our findings are consistent with the predictions of a model that incorporates heterogeneous, private tastes for cooperation. This theory rationalizes our experimental data as it predicts that some variants of the game utilize tastes for cooperation better and can foster cooperation. To our knowledge, our paper is one of only two experimental studies that incorporate private types into their theoretical predictions in an infinitely repeated setting (see Kartal *et al.* (2017) for the other study). Our modeling of tastes for cooperation is closest in spirit to Andreoni and Samuelson (2006), which is an extension of the approach introduced theoretically by the "gang of four," Kreps, Milgrom, Robert, and Wilson.[2,3]

Our paper also relates to the literature on strategic risk/uncertainty. Various authors developed models of strategic risk/uncertainty and others measured its extent and importance in experiments (see, for example, Van Huyck *et al.* (1990, 1991), Morris and Shin (2002), Heinemann *et al.* (2009), Cabrales *et al.* (2010), Dal Bó and Fréchette (2011), Blonski *et al.* (2011), and Andersson *et al.* (2014)). Note that the models in the papers just mentioned can not be applied

[2] Andreoni and Samuelson (2006) analyze theoretically and experimentally a twice-repeated (simultaneous-move) prisoner's dilemma game. Thus, their analysis involves reputation concerns unlike in our setting.

[3] There is an extensive reputation literature dating back to the seminal papers by the "gang of four." Among those, Kreps *et al.* (1982) study a finitely repeated prisoner's dilemma game assuming that each player has a private type that is "committed" to a tit-for-tat strategy with some positive probability and self-interested with the remaining probability. They show that reputation concerns in addition to the presence of conditional cooperators can sustain high levels of cooperation in equilibrium; that is, reputation concerns can motivate self-interested players to imitate a committed type for some time and generate cooperative behavior even with self-interested players. Experimenters have implemented the model by Kreps *et al.* (1982) and variants in the laboratory in order to test its equilibrium predictions (see among others Camerer and Weigelt (1988) and Andreoni and Miller (1993)). While our model involves private information as in those reputation studies, reputation concerns are absent in our setting because imitation does not pay off—this is because neither stakes can increase over time, nor has the game a finite, known endpoint in our setting.

to the analysis of the variants of the prisoner's dilemma game we consider in this paper, while our model can.

The effect of communication has been studied in many contexts and games. Most relevant for our purposes is the potentially cooperation enhancing effect of communication studied in variants of infinitely repeated prisoner's dilemma games with perfect or imperfect monitoring by Embrey *et al.* (2013), Vespa and Wilson (2017), Arechar *et al.* (2017) and Dvořák and Fehrler (2018).[4] Various communication regimes are implemented in these papers. In Embrey *et al.* (2013) and Arechar *et al.* (2017) subjects can only exchange pre-selected messages, while Vespa and Wilson (2017) and Dvořák and Fehrler (2018) allow for free-form chats as in our experiment.[5] In general, communication is reported to increase cooperation rates, which is consistent with the prediction of our model. One exception is Arechar *et al.* (2017) who find that communication (exchange of pre-selected messages in their design) has a negative effect on cooperation when monitoring is noisy and the payoff matrix gives a relatively low return to cooperation.

There are also some experimental studies that implement simultaneous and sequential prisoner's dilemma games. Oskamp (1974) compares cooperation in *finitely repeated* simultaneous and sequential prisoner's dilemma games, whereas Hayashi *et al.* (1999), Ahn *et al.* (2007), and Khadjavi and Lange (2013) run *one-shot* simultaneous and sequential prisoner's dilemma games—the latter also compares behavior of different subject pools (female prisoners versus female university students). Ghidoni and Suetens (2018) is closely related to our study, as they implement simultaneous and sequential infinitely-repeated game versions for all six parameter combinations used in Dal Bó and Fréchette (2011).[6] However, they do not provide a theoretical model or results with communication.

[4] While Embrey *et al.* (2013) and Vespa and Wilson (2017) consider games that resemble prisoner's dilemma games in expectation, Arechar *et al.* (2017) and Dvořák and Fehrler (2017) implement prisoner's dilemma games with noise. In Arechar *et al.* (2017) chosen actions are implemented with errors and in Dvořák and Fehrler (2018) subjects' actions are transformed into signals and a subject's profit in a given round depends on the own action chosen and the *signal* received about the other player's action.

[5] The treatment in Dvořák and Fehrler (2018) with perfect monitoring (where subjects observe both actions and both signals) and free-form communication prior to each round of the repeated game arguably corresponds to our treatment CHAT. Note, however, that their continuation probability of 0.8 is such that cooperation can be supported as a subgame perfect Nash equilibrium in this treatment. This is *not* the case in our treatment CHAT.

[6] Through personal communication, the authors of this paper and Ghidoni and Suetens (2018) became aware of the related research work of the other pair of authors during the time of simultaneous data collection.

3 A Model of Endogenous Strategic Risk

We analyze infinitely repeated prisoner's dilemma games in which (i) players have a taste for cooperation, (ii) tastes are heterogeneous, and (iii) each player's taste for cooperation is private information. Consider the following payoff matrix for a symmetric prisoner's dilemma stage game:

$$\begin{array}{c|c|c|} & C & D \\ \hline C & c,\ c & a,\ b \\ \hline D & b,\ a & d,\ d \\ \hline \end{array}.$$

The payoff parameters a, b, c, and d are such that $b > c > d > a$ and $2c > b + a$. Let $\Gamma(\delta)$ denote an arbitrary symmetric infinitely repeated prisoner's dilemma game with discount factor δ, and let $\Gamma(a, b, c, d, \delta)$ denote a symmetric infinitely repeated game with discount factor δ and monetary payoffs a, b, c, and d as in the table above.

At the beginning of each period $t = 0, 1, \ldots$, two players make a choice between C or D. We model preferences and specify *stage-game utilities* in game $\Gamma(a, b, c, d, \delta)$ as follows. Player i receives a utility of $u(a_i, a_j, \gamma_i)$ from choosing $a_i \in \{C, D\}$, where $a_j \in \{C, D\}$ denotes player j's choice in the stage game, and γ_i represents player i's taste for cooperation, which is i's private information. We interpret γ as a measure of taste for (conditional) cooperation in a sense that will become clear below (in particular, the higher the value of γ_i for player i, the stronger the taste of i for cooperation).

Stage-game utility $u(a_i, a_j, \gamma_i)$ is continuous in γ_i, and γ_i is an independently and identically distributed draw from a commonly known distribution $F(\gamma)$ with a continuous density on $[\underline{\gamma}, \bar{\gamma}]$. Stage-game utilities are such that $u(a_i, C, \gamma_i) > u(a_i, D, \gamma_i)$ for $a_i \in \{C, D\}$; that is, i prefers j to cooperate regardless of i's choice as in the standard prisoner's dilemma game. Moreover, as in the standard game $u(D, D, \gamma_i) > u(C, D, \gamma_i)$ for $\gamma_i < \bar{\gamma}$.[7] We assume that $\partial u(C, C, \gamma_i)/\partial c > 0$, $\partial u(C, D, \gamma_i)/\partial a > 0$, $\partial u(D, C, \gamma_i)/\partial b > 0$, and $\partial u(D, D, \gamma_i)/\partial d > 0$. That is, all else equal an increase in the monetary payoff to i resulting from (a_i, a_j) increases the stage-game utility to i from (a_i, a_j).

To make the analysis tractable, we adopt and extend the approach in Dal Bó and Fréchette (2011) and consider subjects contemplating whether to play a (conditionally) cooperative strategy

[7] Note that comparative static results are not affected if $u(D, D, \gamma_i) \leq u(C, D, \gamma_i)$ for large γ_i; such an assumption would be consistent with, for example, the presence of some unconditional cooperators among players and require us to consider a third strategy, namely the "always cooperate" strategy, which would not affect our main conclusions.

or the "always defect" strategy (hereafter, AD). The (conditionally) cooperative strategy (hereafter, CS) is a strategy such as grim, tit-for-tat or limited punishment that starts by cooperating and depends on the past behavior of players.[8] Despite its simplicity, the model generates interesting and novel comparative predictions regarding cooperative behavior, which is borne out by the data. Moreover, this simplicity in focus implies that all the results extend to symmetric continuous-action dilemma games (for example, Cournot competition), once we determine the socially efficient (*i.e.*, the most cooperative) payoff, the temptation and sucker payoffs as well as the stage game Nash equilibrium payoff, which is strictly dominated by the efficient payoff in dilemma games.

Consider an infinitely repeated prisoner's dilemma game and let

$$\begin{aligned}\Pi(\delta, p, \gamma) \;&=\; \left[p\frac{u(C,C,\gamma)}{1-\delta} + (1-p)\left(u(C,D,\gamma) + \delta\frac{u(D,D,\gamma)}{1-\delta}\right)\right] - \\ &\left[p\left(u(D,C,\gamma) + \delta\frac{u(D,D,\gamma)}{1-\delta}\right) + (1-p)\frac{u(D,D,\gamma)}{1-\delta}\right].\end{aligned}$$

The term $\Pi(\delta, p, \gamma)$ denotes the *net* expected benefit to a type-γ player from choosing CS (relative to AD) if a fraction p of subjects chooses CS and the rest chooses AD. This is true because the first term and the second term inside the square brackets on the right-hand side equal the respective expected payoff to a type-γ player from CS and AD under the assumption that the opponent plays CS with probability p and AD with the remaining probability. The terms p and $\Pi(\delta, p, \gamma)$ relate to the axiomatic approach to equilibrium selection in Blonski *et al.* (2011) who emphasize the importance of the "sucker payoff" as a determinant of cooperation in repeated games and the "basin of attraction" of AD (relative to CS) elaborated on in Dal Bó and Fréchette (2011). In our model, an increase in p decreases the basin of attraction of AD and an increase in the sucker payoff increases $\Pi(\delta, p, \gamma)$. We extend the previous literature endogenizing p, which represents the strategic risk of cooperation, as the fraction of conditional cooperators is determined endogenously in equilibrium and pins down the value of p.

In our model, p is less than one in equilibrium as we are interested in understanding cooperation in games in which a nontrivial subset of γ-types are faced with a dilemma because δ is away from one, b is not close to c, c is not substantially higher than d, *etc.*[9] To fix ideas, however, we first consider a setting in which $p = 1$ so that there is no strategic risk. According to assumption

[8] Whether the cooperative strategy is grim, tit-for-tat or limited punishment is inconsequential given our simplifying assumption that players select into CS or AD strategies at the beginning of the game,

[9] As an example, in the six repeated games analyzed by Dal Bó and Fréchette (2011), only one game comes close to generating full cooperation, and in that game, b and c are very close.

A1 below, in the absence of strategic risk, there exists a type-γ_1 player who strictly prefers CS over AD, but there also exists a type-γ_2 player who strictly prefers AD.

Assumption A1 Assume that δ is bounded above away from one. There exists γ_1 and γ_2 such that $\underline{\gamma} \leq \gamma_2 < \gamma_1 \leq \bar{\gamma}$, $\Pi(\delta, 1, \gamma_1) > 0$ and $\Pi(\delta, 1, \gamma_2) < 0$.

To understand A1 better, consider the simple case where $\delta = 0$ so that the game is one shot. If $\Pi(0, 1, \gamma_1) > 0$, then this implies that in a one-shot prisoner's dilemma game, a type-γ_1 player finds it optimal to cooperate against an opponent who will certainly cooperate (because we set $p = 1$), whereas $\Pi(0, 1, \gamma_2) < 0$ implies that type-γ_2 player finds it optimal to defect in that case. The interpretation is that type-γ_1 is a (conditional) cooperator and has a strong taste for mutual cooperation, but there are other types who do not share those preferences such as a selfish type with standard preferences, a spiteful type, or a type with a weak taste for cooperation.

Even a type-γ_1 player for whom $\Pi(\delta, 1, \gamma_1) > 0$ could have $\Pi(\delta, p, \gamma_1) < 0$ with $p < 1$ and thus prefer defection. In less formal terms, there will naturally be types who strictly prefer cooperating against an opponent who will certainly cooperate but will defect if they are not sufficiently confident that the opponent will cooperate because the risk of obtaining the sucker payoff kicks in. As a result, A1 is too weak to ensure that there is (at least some) cooperation in equilibrium, and there are various ways to make it stronger. We choose one such possibility below. The following assumption strengthens the condition regarding γ_1 in A1 (hence the name A1S) and keeps the condition regarding γ_2 as in A1.

Assumption A1S Assume that δ is bounded above away from one. There exists γ_1 and γ_2 such that $\underline{\gamma} \leq \gamma_2 < \gamma_1 \leq \bar{\gamma}$, $\Pi(\delta, 1 - F(\gamma_1), \gamma_1) > 0$ and $\Pi(\delta, 1, \gamma_2) < 0$.

The term $\Pi(\delta, 1 - F(\gamma_1), \gamma_1)$ in A1S is the *net* expected benefit of CS to a type-γ_1 player assuming that the opponent chooses CS if her type is greater than γ_1 and AD otherwise. Thus, A1S implies that there exists a γ_1-type that strictly prefers cooperation if the strategic risk of cooperation equals $1 - F(\gamma_1)$. Assumption A2 below imposes a *monotonicity* condition on $\Pi(\delta, p, \gamma)$ for tractability, and implies that if a γ_1-type strictly prefers CS over AD with a strategic risk equal to $1 - F(\gamma_1)$ as postulated in A1S, then any type greater than γ_1 strictly prefers CS, as well, because the strength of preferences for cooperation increases in γ.

Assumption A2 If $\gamma' > \gamma$, then $\Pi(\delta, p, \gamma') > \Pi(\delta, p, \gamma)$.

In all of our results below, we assume that A1S and A2 hold. We now provide some simple examples in which A1S and A2 hold.

Example 1 Consider a stage game payoff matrix with $a = 12$, $b = 50$, $c = 32$, and $d = 25$ (this is the payoff matrix that we use in our experimental design). Assume that $\delta = 0.5$. Consider preferences that are represented by the following utility specification: $u(C, C, \gamma_i) = c^{\gamma_i}$ where $\gamma_i > 0$, and for every other (a_i, a_j) pair, stage-game utilities equal stage-game monetary payoffs. It can easily be checked that A2 is satisfied. If, for example, γ_i is drawn from a uniform distribution between 1 and 1.25, then $\gamma_1 = 1.2$ and $\gamma_2 = 1$ satisfy A1S.

Example 2 Consider the stage game payoff matrix in Example 1 and assume that $\delta = 0.8$. Consider preferences that give rise to the following utility specification: $u(C, C, \gamma_i) = c + \gamma_i$, $u(C, D, \gamma_i) = a - \alpha(\gamma_i)$, $u(D, C, \gamma_i) = b - \beta(\gamma_i)$, and finally, $u(D, D, \gamma_i) = d$. Assume for example that $\alpha(\gamma_i) = |\gamma_i|$ and $\beta(\gamma_i) = \gamma_i$. Then, A2 is satisfied. If γ_i is drawn from a normal distribution with mean 0 and sufficiently large σ, then A1S is also satisfied (for example, $\gamma_1 = 2$ and $\gamma_2 = -2$ satisfy A1S if σ equals 10).[10]

Assumptions A1S and A2 imply that in *every* game there are at least some (conditional) cooperators who will optimally choose a cooperative strategy—either because the strategic risk of cooperation is not too high as there are sufficiently many of them (*i.e.*, $1 - F(\gamma_1)$ is high enough), or because their taste for mutual cooperation is high enough to compensate for the strategic risk of being the sucker ((*i.e.*, γ_1 is high enough). These assumptions enable us to endogenize strategic risk in the analysis of repeated prisoner's dilemma games and generate interesting comparative static results.

We next consider the perfect Bayesian Nash equilibria in our benchmark case, the infinitely repeated simultaneous prisoner's dilemma game. To justify our interest in mutual cooperation in game $\Gamma(\delta)$, and to have consistency with the inequality $2c > b + a$ stated above, we assume that $u(C, C, \gamma_i) + u(C, C, \gamma_j) > u(C, D, \gamma_i) + u(D, C, \gamma_j)$ for (at least) γ_i values for which $\Pi(\delta, 1, \gamma_i) \geq 0$ holds. As stated below in Proposition 1, there always exists an equilibrium cutoff type $\gamma^* \in (\underline{\gamma}, \bar{\gamma})$ such that $\Pi(\delta, 1 - F(\gamma^*), \gamma^*) = 0$, and a type-$\gamma$ player selects CS if $\gamma > \gamma^*$ and AD otherwise. (Note that the detailed proofs of our Propositions are provided in the Appendix.) Moreover, we show that every equilibrium must be symmetric. It is not possible to claim that there is a unique equilibrium cutoff without making further assumptions. This is mostly inconsequential for our main results in Proposition 2. For Propositions 1 and 3, we make statements focusing on the most cooperative equilibrium; that is, we focus on the lowest equilibrium cutoff γ^* in the game.

[10] It is also possible to generate examples in which A1S and A2 hold with Charness and Rabin (2002) or Fehr and Schmidt (1999) preferences.

Strategic risk arises endogenously in this model since the expected fraction of cooperators, which equals $1 - F(\gamma^*)$, is an equilibrium object. As a result, a change in a stage game payoff, parameter or the game format has a direct effect on γ^*, and hence the cooperation rate as stated in Propositions 1-3 below. Note that for part (ii) of Proposition 1 we need the intuitive assumption that if games $\Gamma(\delta)$ and $\Gamma'(\delta)$ give the same monetary payoff to i from (a_i, a_j), then $u(a_i, a_j, \gamma_i)$ is identical in games $\Gamma(\delta)$ and $\Gamma'(\delta)$.

Proposition 1 ***(i)*** *There exists a Bayesian Nash equilibrium in game* $\Gamma(a, b, c, d, \delta)$. *Every equilibrium is symmetric and consists of a cutoff* γ^* *such that a player with type* $\gamma > \gamma^*$ $(\gamma < \gamma^*)$ *chooses CS (AD).* ***(ii)*** *If* $a' < a$ $(b' > b)$, *then the cooperation rate in game* $\Gamma(a', b, c, d, \delta)$ $(\Gamma(a, b', c, d, \delta))$ *is strictly lower than the cooperation rate in game* $\Gamma(a, b, c, d, \delta)$. ***(iii)*** *If* $\delta' < \delta$, *then the cooperation rate in game* $\Gamma(a, b, c, d, \delta')$ *is strictly lower than the cooperation rate in game* $\Gamma(a, b, c, d, \delta)$. ***(iv)*** *If* $F'(\gamma)$ *first order stochastically dominates* $F(\gamma)$, *then a population characterized by* $F'(\gamma)$ *induces a higher cooperation rate than one characterized by* $F(\gamma)$.

Among other things, Proposition 1 implies that a decrease in the sucker payoff a (all else equal) increases the equilibrium cutoff type and reduces the equilibrium cooperation rate. It also implies that an increase in δ reduces the equilibrium cutoff type and increases the cooperation rate. The last statement of the proposition, which concerns the distribution of preferences is particularly relevant for experimental data and subject pool effects as we discuss in more detail in Section 4 (among others Fehr *et al.* (2006) and Engelmann and Normann (2010) find significant subject pool effects in their experiments).

We next analyze how preferences for cooperation influence cooperation and coordination on cooperation under different infinitely repeated institutions. We start with the comparison of the infinitely repeated simultaneous prisoner's dilemma game to the infinitely repeated sequential prisoner's dilemma game. The stage game payoffs and parameters are identical across the two games, but players make choices simultaneously in the stage game of one variant, and sequentially in the other. Note that standard theory makes identical equilibrium predictions for the two games, but our predictions are different as seen in Proposition 2 below.

Proposition 2 ***(i)*** *The first-mover cooperation rate in the sequential game is strictly higher than the cooperation rate in the simultaneous game.* ***(ii)*** *The second-mover cooperation rate conditional on the first-mover choosing* C *in the initial period is higher than the cooperation rate in the simultaneous game and the first mover cooperation rate in the sequential game.* ***(iii)*** *The rate of* *coordination on cooperation and the continuation game cooperation rate are strictly higher in the sequential game.*

The intuition of Proposition 2 is based on the simple idea that sequential moves reduce strategic risk for both the first mover and the second mover, and the reduction in strategic risk is associated with a boost in the incentive to cooperate. To see this, note first that there is no strategic risk from the viewpoint of a second mover who observes cooperation by the first mover. Thus, any type-γ second mover for whom $\Pi(\delta, 1, \gamma) > 0$ holds strictly prefers the cooperative strategy once the first mover chooses to cooperate. As a result, the cooperation rate of second movers (conditional on first-mover cooperation) is determined by $\Pi(\delta, 1, \gamma_2^*) = 0$. It follows that $\gamma_2^* < \gamma^*$ because cooperation is risky in the simultaneous game, unlike the situation for a second mover after the first mover cooperates, and the increased strategic risk reduces the incentive to cooperate. In turn, the first mover in the sequential game expects a strictly higher cooperation rate than the cooperation rate a player can expect in the simultaneous game; *i.e.*, $1-F(\gamma_2^*) > 1-F(\gamma^*)$. Thus, the strategic risk for the first mover is also reduced, and $\gamma_1^* < \gamma^*$ must hold as well. However, strategic risk in the sequential game is higher for the first mover than the second mover, and as a result $\gamma_2^* < \gamma_1^*$.

Note that we cannot prove that the expected first period cooperation rate is higher in the sequential game than in the simultaneous game simply because cooperation in the sequential game has a correlated structure even in the first period. Thus, even though we know that $\gamma_1^* < \gamma^*$ and that $\gamma_2^* < \gamma^*$, we cannot know how the unconditional probability that the second mover cooperates in the first period of the sequential game (this probability is equal to $(1 - F(\gamma_1^*))(1 - F(\gamma_2^*))$ compares to the predicted first period cooperation rate of the simultaneous game (this is equal to $1 - F(\gamma^*)$). Intuitively, the unconditional probability that a second mover cooperates may be relatively low since some second movers who would choose to cooperate in the simultaneous game will meet first movers who choose to defect and will subsequently defect themselves. However, as the rate of coordination on cooperation is strictly higher in the initial period of the sequential game than in the simultaneous game, the expected continuation game cooperation rate is strictly higher in the sequential game. Finally, note that incorporating preferences for reciprocity into the sequential move game would make our results only stronger. That is, one may naturally assume that observing the first mover choosing C may result in a preference distribution $F'(\gamma)$ for second movers such that $F'(\gamma)$ first order stochastically dominates $F(\gamma)$. This reciprocity effect reduces γ_2^*. In turn, the reduction in γ_2^* reduces γ_1^*.

Next, we consider the effect of communication (*i.e.*, free-flow chat) on cooperation. We start with the following assumption. Players who intend to use a cooperative strategy in (the most cooperative) equilibrium will always communicate their intention to cooperate, and will indeed choose the cooperative strategy upon mutual agreement and the defecting strategy otherwise. We also assume that a positive fraction of players who will choose AD are nevertheless honest and will signal their intention truthfully (via communication or refusing to communicate). This is a reasonable assumption given what we already know from experimental studies on honesty and communication. Lying has a psychological cost to many people, and indeed, a large majority of communication is found to be truthful in many experiments (see among others Embrey *et al.* (2013), Arechar *et al.* (2017) and Vespa and Wilson (2017) who explore infinitely repeated interactions with communication). This is true in our experiment as well. In what follows, "honesty rate" refers to the fraction of players who will choose AD but will not lie about their intentions.

Proposition 3 ***(i)*** *With any strictly positive honesty rate, the rate of coordination on cooperation and the continuation game cooperation rate are strictly higher in the communication game than in the simultaneous game.* ***(ii)*** *If the honesty rate is sufficiently high, then the rate of coordination on cooperation and the continuation game cooperation rate are strictly higher in the communication game than in the sequential game.*

For intuition, consider the simpler case in which there is no dishonesty among players. Given our assumptions on communication, the most cooperative equilibrium has a correlated structure (as in the sequential game) and a type-γ player for whom $\Pi(\delta, 1, \gamma) \geq 0$ holds will choose to cooperate *if and only if* the player meets and communicates with a type-γ' player for whom $\Pi(\delta, 1, \gamma') \geq 0$ also holds. Thus, the equilibrium cutoff in the communication game is given by $\Pi(\delta, 1, \gamma_2^*) = 0$, where γ_2^* is precisely the equilibrium cutoff for the second mover conditional on first mover cooperation in the sequential game. It follows that the coordination rate in the communication game is equal to $(1-F(\gamma_2^*))^2$, which also equals the first round cooperation rate because the decision to cooperate is perfectly coordinated without any dishonesty in communication. As explained before while comparing the sequential game to the simultaneous game, $\gamma_2^* < \gamma^*$ and $\gamma_2^* < \gamma_1^*$. Thus, the coordination rate in the communication game is strictly greater than $(1 - F(\gamma^*))^2$ and $(1-F(\gamma_1^*))(1-F(\gamma_2^*))$, the respective coordination rate in the simultaneous game and the sequential game. These arguments extend to the case in which the honesty rate in communication is strictly positive (this is relevant for the comparison to the simultaneous game) or sufficiently high (this is relevant for the comparison to the sequential game).[11]

Preferences for reciprocity may also be evoked in this game (as in the sequential game). A player may feel reciprocity towards an opponent who proposes mutual cooperation or declares the intention to cooperate. In fact, reciprocity may influence behavior even more in the communication game than in the sequential game, because players likely feel less anonymous in the presence of a chat opportunity. This would then result in a preference distribution $F'(\gamma)$ for recipients of "nice communication such that $F'(\gamma)$ first order stochastically dominates $F(\gamma)$, which increases equilibrium cooperation and coordination on cooperation further. Also, note that the first-period cooperation rate may be higher in the communication game than in the simultaneous game especially if the simultaneous game is one with high strategic risk. This would mean that $1 - F(\gamma^*)$ is very low and, thus, it may be easier for $(1 - F(\gamma_2^*))^2$ to exceed $1 - F(\gamma^*)$. As discussed in more detail below in Section 3.2, we choose a game with high strategic risk for our experimental treatments, which has been analyzed by Dal Bó and Fréchette (2011) and shown to give rise to very low cooperation rates in the benchmark simultaneous setting.

Apart from the variations we presented above, our model can be extended in other ways. For example, it can be extended to noisy prisoner's dilemma games (e.g. introducing imperfect monitoring to the benchmark simultaneous stage game should conceivably reduce cooperation in the repeated game). However, a formal analysis of noisy games with endogenous strategic risk is beyond the scope of our current work and an avenue for future research.

4 Treatments, Hypotheses and Protocols

4.1 Treatments and Hypotheses

Our experiment consists of three treatments. In all treatments we implemented $\delta = 0.5$ and used the following game matrix (corresponding to the treatment with $R = 32$ and $\delta = 0.5$ in Dal Bó and Fréchette (2011), henceforth referred to as DB&F):

	C	D
C	32, 32	12, 50
D	50, 12	25, 25

[11] We cannot claim that the first-period cooperation rate is higher in the communication game than in the simultaneous game or the sequential game (e.g., the first-period cooperation rate equals $(1 - F(\gamma_2^*))^2$ in the communication game without dishonesty, whereas it equals $1 - F(\gamma^*)$ in the simultaneous game).

That is, in view of the general matrix introduced in Section 3, $a = 12$, $b = 50$, $c = 32$, $d = 25$.

In this game, theory uniquely predicts defection under the assumption that players are exclusively motivated by their own monetary payoffs. DB&F found that the game gave rise to very low cooperation rates especially after subjects gained experience. We chose this game as our baseline repeated simultaneous game because there is arguably more room for variation in behavior across treatments if we implement a baseline which is experimentally shown to have high strategic risk and a low cooperation rate.

In what follows, we use the terms "match" and "indefinitely repeated game" interchangeably. In the baseline treatment (denoted by "SIM"), players make their choices simultaneously in every round of a match as in DB&F, whereas in the second treatment (denoted by "SEQ"), players make their choices sequentially; *i.e.*, the stage game is a sequential prisoner's dilemma game, and everything else is the same as in the SIM treatment. In the third treatment (denoted by "CHAT"), everything is exactly the same as in the SIM treatment with one exception; players have the chance to chat in free form at the beginning of every round of a match. In our theory, whether communication takes place once at the beginning of the first round, or at the beginning of every round does not matter, barring the *reciprocity* effect of communication on cooperation, which we discussed after Proposition 3. We conjectured that such a reciprocity effect is easier to maintain if pairs can repeatedly communicate. Therefore, in our experimental design we chose to allow for communication at the beginning of every round.

We now state our hypotheses based on Propositions 2 and 3, as well as our discussion after Proposition 3. Let λ_{SIM} and λ_{CHAT} denote the cooperation rate in the first round of a match in the SIM and CHAT treatments, respectively. For the SEQ treatment, λ^1_{SEQ} and λ^2_{SEQ} denote the respective first mover and second mover cooperation rate in the first round, the latter being conditional on the first mover choosing C. In formal terms, let γ_{SIM} and γ_{CHAT} denote the respective equilibrium cutoff in the SIM treatment and the CHAT treatment. Also, let γ^1_{SEQ} and γ^2_{SEQ} denote the respective equilibrium cutoff for the first mover and the second mover (conditional on first mover cooperation) in the SEQ treatment. Then, $\lambda_{SIM} = 1 - F(\gamma_{SIM})$, $\lambda^1_{SEQ} = 1 - F(\gamma^1_{SEQ})$, and $\lambda^2_{SEQ} = 1 - F(\gamma^2_{SEQ})$. Next, note that λ_{CHAT} is such that $\lambda_{CHAT} \in \left((1 - F(\gamma_{CHAT}))^2, 1 - F(\gamma_{CHAT})\right)$ as the exact cooperation rate depends on the honesty rate in communication. If, for example, the honesty rate is 100 percent, then cooperation is perfectly coordinated and thus λ_{CHAT} equals $(1 - F(\gamma_{CHAT}))^2$, whereas if the honesty rate is 0, then λ_{CHAT} equals $1 - F(\gamma_{CHAT})$. Of course, γ_{CHAT} also varies depending on the honesty rate; in particular, the higher the honesty rate, the lower the cutoff γ_{CHAT}. Note that part (ii) in Hypothesis 1 is based on our discussion following Proposition 3 since the SIM treatment game has high strategic risk; that is, the equilibrium cutoff type is high and should be close to $\bar{\gamma}$ as DB&F found very low cooperation rates in this game.

Hypothesis 1 $\lambda^2_{SEQ} > \lambda^1_{SEQ} > \lambda_{SIM}$: *The first-mover cooperation rate in the first round of a match in SEQ is strictly higher than the cooperation rate of a player in the first period of a match in SIM and strictly lower than the second-mover cooperation rate in the first round of a match in SEQ (conditional on first mover cooperation). These statements also hold on average across all rounds.*[12] **(ii)** $\lambda_{CHAT} > \lambda_{SIM}$: *Since strategic risk is high in the SIM treatment, the cooperation rate of a player in the first period of a match in CHAT is higher than that in SIM. The statement also holds on average across all rounds. That is, the average cooperation rate in CHAT is strictly higher than the average cooperation rate in SIM.*

The next hypothesis concerns coordination on cooperation and follows from Proposition 3. Note that in the CHAT treatment, the coordination rate is given by $(1 - F(\gamma_{CHAT}))^2$, whereas it equals $\lambda^1_{SEQ} \times \lambda^2_{SEQ}$ and λ^2_{SIM} in the SIM and SEQ treatments, respectively.

Hypothesis 2 $(1-F(\gamma_{CHAT}))^2 > \lambda^1_{SEQ} \times \lambda^2_{SEQ} > (\lambda_{SIM})^2$: *The rate of coordination on cooperation (and the continuation game cooperation rate) is highest in the CHAT treatment, followed by the SEQ treatment and lowest in the SIM treatment.*

4.2 Experimental Protocol

Our design is between-subject, *i.e.*, each subject participated in only one treatment. For each treatment, we have six independent matching groups. In each treatment, each subject participated in a sequence of infinitely repeated games. We refer to each indefinitely repeated game as a match and each repetition of the game within a match as a round.

The implementation of the SIM treatment was very similar to that in DB&F for comparability. At the beginning of each session in the SEQ treatment, each subject was randomly assigned to be a first mover or a second mover. Subjects remained in the same role throughout the session. Each session of the experiment ended after the first match that was completed after 75 minutes had passed, or after 60 matches had been completed, whichever happened sooner. Subjects were informed about this.

[12] That is, the average first-mover cooperation rate in SEQ is strictly higher than the average cooperation rate of a player in SIM and is strictly lower than the average second-mover cooperation rate in SEQ (conditional on first mover cooperation).

We implemented an infinitely repeated game in the lab by using a random continuation rule. The probability of continuation after each round was equal to $\delta = 0.5$ in each treatment. The experiment was conducted at the experimental laboratory of the Vienna Center for Experimental Economics (VCEE) at the University of Vienna. Subjects were recruited from the subject pool maintained at the VCEE via ORSEE (Greiner, 2015). A total of 280 subjects participated in our experiment. All sessions were conducted through computer terminals, using a program written in z-Tree (Fischbacher, 2007).[13] At the end of each session, the total number of points earned by each subject was converted to Euros at the exchange rate of €0.007 for every point scored, and paid privately in cash. The average payoff per subject was €19.43. The instructions for the treatments can be found in the Online Appendix.

In all matching groups of treatment SIM, 60 repeated games were played (see Table 8 in the Online Appendix). However, the sequential version of the prisoner's dilemma game and especially the one combined with chat took, as expected, longer to complete than a simple prisoner's dilemma game. Hence, fewer than the maximum number of 60 repeated games were played in treatment SEQ and CHAT. The results mentioned in footnote 15 and elaborated on in Section C in the Online Appendix suggest that our main results would not have changed had more repeated games been played.

5 Experimental results

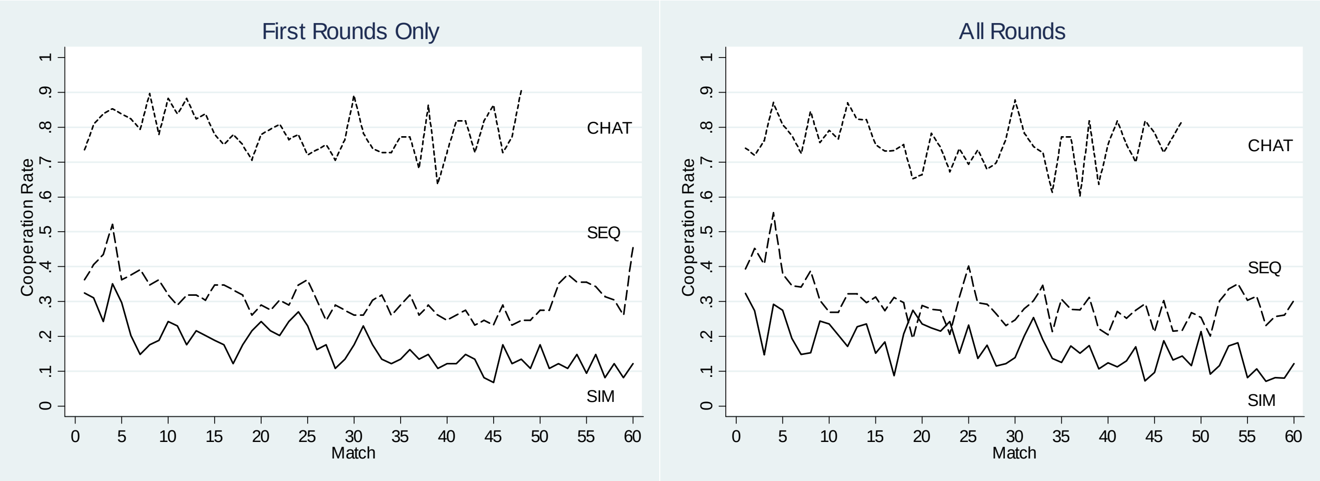
We report the results of our experiment in three parts. First, we analyze cooperation and coordination on cooperation across the three treatments and test for statistical differences. Second, we analyze the nature and content of the chats in treatment CHAT and relate them to the choices subjects made. Finally, using the structural frequency estimation method (SFEM) developed in DB&F, we estimate the types and shares of strategies used in the three treatments and, where appropriate, test for statistical differences across the three treatments.

5.1 Cooperation and mutual cooperation rates

Figure 1 shows the evolution of the average cooperation rate in treatments SIM and CHAT and the first-mover cooperation rate in treatment SEQ. In particular, we show the averages of only first rounds of all matches on the left and averages of all rounds of all matches on the right in

[13] We used adapted versions of a z-Tree program developed by Sevgi Yuksel and Emanuel Vespa for the papers Fréchette and Yuksel (2017) and Vespa (2015).

Figure 1: Evolution of Cooperation in Treatments



Notes: Only first-mover data for treatment SEQ. The left (right) panel uses data from first (all) rounds of all matches.

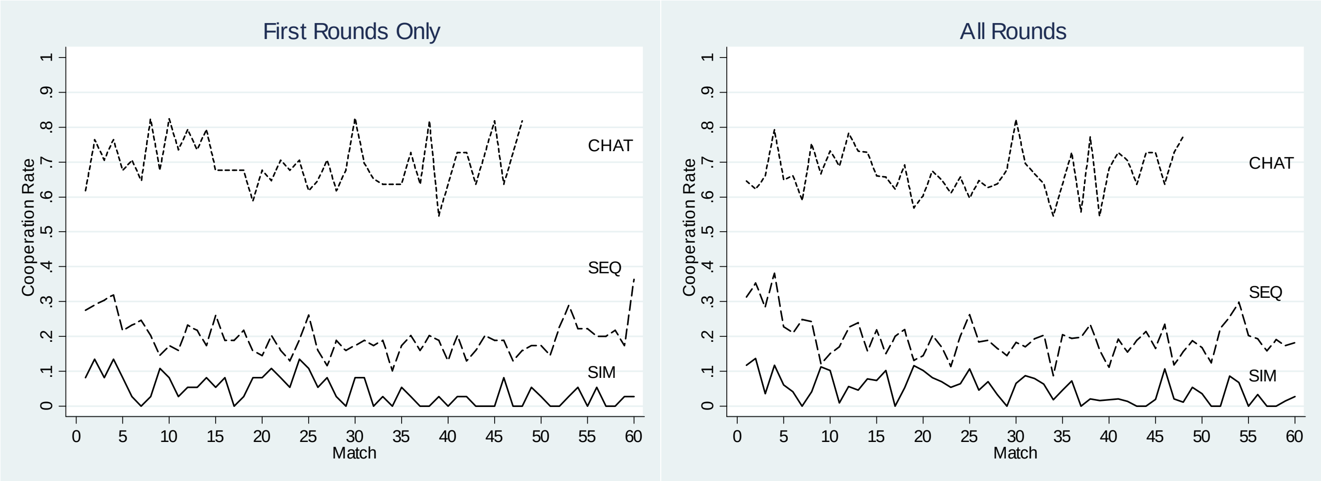
Figure 1. We observe a clearly higher cooperation rate in SEQ in comparison to SIM, and a much higher cooperation rate in CHAT than in SIM and SEQ (first movers only). Furthermore, average cooperation rates do not appear to differ much depending on whether only first rounds or all rounds of matches are considered.

Figure 3 in the Online Appendix shows the evolution of the average cooperation rate of all matching groups in SIM, CHAT, and SEQ (first movers only). Figure 3 reveals that there is quite some heterogeneity across the various matching groups within each treatment.

Figures 1 and 3 show only the first-mover behavior in SEQ. To also account for second-mover behavior in SEQ, Figure 2 shows the evolution of average mutual-cooperation rates, where mutual cooperation is the outcome in which both players coordinate on cooperation. The differences in average cooperation rates shown in Figure 1, carry over to mutual-cooperation rates, with these rates being clearly higher in SEQ than in SIM, and again much higher in CHAT than in SIM and SEQ.

Table 1 shows summary statistics of cooperation rates in all three treatments along with the results of statistical tests (indicated by "<" and ">" signs). The information contained in this table allows for a straightforward comparison of our results in treatment SIM with the results of the corresponding treatment in DB&F (p. 417, Table 3, $\delta = 1/2$, $R = 32$). We find that behavior in the first repeated game is quite similar in these two studies. Indeed, the average cooperation rate

Figure 2: Evolution of Cooperatve Outcomes in Treatments



Note: The left (right) panel uses data from first (all) rounds of all matches.

in the first round (in all rounds) of the first repeated game of SIM is 32.42 (32.35), while it is 34.09 (28.33) in DB&F. However, if all repeated games are considered, clear differences between SIM and the corresponding treatment in DB&F emerge. While the average cooperation rate in first rounds (in all rounds) of all matches is 17.05 (16.79) in SIM, it is just 9.81 (9.82) in DB&F. The higher cooperation rates in our treatment appear to be due to the aforementioned higher heterogeneity of observed cooperation rates across our six matching groups.[14]

To formally assess across-treatment differences in our data, we run probit regressions of cooperation and mutual-cooperation rates on a treatment dummy and include data from pairs of treatments, clustering observations at the matching group level. In order to economize on space, we focus on the lower part of Table 1, which shows summary statistics of cooperation rates and test results using data from *all* matches.

The differences in cooperation rates between SEQ (first movers) and SIM are statistically significant at the 5 percent level (regardless of whether data comes from only first rounds or all rounds of all matches). Moreover, the across-treatment differences in treatments CHAT and SIM are highly significantly different—we observe the same when we compare CHAT and SEQ (first movers).

Table 1 also shows the average second-mover cooperation rate in SEQ after the first mover chooses C or D. Using similar probit regressions as above, we tested behavior of second movers in

[14] The higher cooperation rate in our treatment SIM than in the corresponding treatment in DB&F can be interpreted as a subject pool effect which is in line with Proposition 1(iv).

SEQ after observing that the first mover chose C against (a) second mover behavior after observing that the first mover chose D, (b) first mover behavior in treatment SEQ, and behavior in (c) treatment SIM, and (d) treatment CHAT. The test results are indicated in Table 1, and results for (b), (c) and (d) use the notation λ_{SIM}, λ_{CHAT}, λ^1_{SEQ}, and λ^2_{SEQ}, as introduced in Section 3.1, page 15. Taking all matches into account, and for both first and all rounds, the second-mover cooperation rate in SEQ after observing that the first mover chose C is significantly higher than the first-mover cooperation rate in SEQ and the cooperation rate in SIM. It is, however, lower than the cooperation rate in CHAT if first rounds of all matches are considered.

Table 2 shows summary statistics and test results regarding coordination on cooperation. To illustrate, the average mutual-cooperation rate in all rounds of all matches in SIM, SEQ and CHAT is 4.99, 19.92, and 66.37 percent, respectively. The test results shown in Table 2 indicate that all pairwise comparisons of mutual-cooperation rates are statistically significant at the 1 percent level, and the signs are as predicted.[15]

The test results indicated in Tables 1 and 2 confirm Hypotheses 1 and 2. To be more precise, we state the following experimental results, which hold for average behavior in first rounds as well as in all rounds of each match:

Experimental Result 1: *The first-mover cooperation rate in SEQ is strictly higher than the cooperation rate in SIM, and strictly lower than the second-mover cooperation rate in SEQ (conditional on first mover cooperation). The cooperation rate in CHAT is strictly higher than the cooperation rate in SIM.*[16]

Experimental Result 2: *The rate of coordination on cooperation is the highest in CHAT, followed by SEQ and the lowest in SIM.*

5.2 Detailed analysis of treatment CHAT

Recall that in treatment CHAT, in each round of a match subjects could exchange free-form chat messages with their opponent prior to making their choices. In this section, we first briefly analyze

[15] One might wonder what levels the cooperation and mutual cooperation rates reported in Tables 1 and 2 would have converged to had the experiment been conducted over a very long time horizon. In Online Appendix C we present an analysis that uses techniques proposed in Noussair *et al.* (1995) or Barut *et al.* (2002) that suggest that none of the reported (mutual) cooperation rates would have converged to zero in the long run. Moreover, and more important for our purposes, the (mutual) cooperation rates in both treatments SEQ and CHAT would have stayed significantly higher that in those in treatment SIM.

[16] The cooperation rate in CHAT is also higher than the first-mover (second-mover) cooperation rate in SEQ (conditional on first mover cooperation). However, we did not have a clear prediction for this comparison.

Table 1: Summary statistics and test results: Cooperation rates

First repeated game

First round					All rounds				
SEQ		**SIM**		**CHAT**	**SEQ**		**SIM**		**CHAT**
1st Mover					1st Mover				
36.23	$>$	32.43	$<^{***}$	73.53	39.30	$>$	32.35	$<^{***}$	74.05
SEQ					**SEQ**				
2nd Mover					2nd Mover				
After					After				
C D					C D				
76.00 $>^{***}$ 11.36					79.75 $>^{***}$ 7.38				
$\lambda^2_{SEQ} >^{***} \lambda_{SIM}$, $\lambda^2_{SEQ} > \lambda_{CHAT}$					$\lambda^2_{SEQ} >^{***} \lambda_{SIM}$, $\lambda^2_{SEQ} > \lambda_{CHAT}$				
$\lambda^2_{SEQ} >^{***} \lambda^1_{SEQ}$, $\lambda^1_{SEQ} <^{***} \lambda_{CHAT}$					$\lambda^2_{SEQ} >^{***} \lambda^1_{SEQ}$, $\lambda^1_{SEQ} <^{***} \lambda_{CHAT}$				

All repeated games

First round					All rounds				
SEQ		**SIM**		**CHAT**	**SEQ**		**SIM**		**CHAT**
1st Mover					1st Mover				
30.90	$>^{**}$	17.05	$<^{***}$	79.08	29.63	$>^{**}$	16.79	$<^{***}$	74.59
SEQ					**SEQ**				
2nd Mover					2nd Mover				
After					After				
C D					C D				
62.31 $>^{***}$ 3.32					67.21 $>^{***}$ 7.38				
$\lambda^2_{SEQ} >^{***} \lambda_{SIM}$, $\lambda^2_{SEQ} <^{***} \lambda_{CHAT}$					$\lambda^2_{SEQ} >^{***} \lambda_{SIM}$, $\lambda^2_{SEQ} < \lambda_{CHAT}$				
$\lambda^2_{SEQ} >^{***} \lambda^1_{SEQ}$, $\lambda^1_{SEQ} <^{***} \lambda_{CHAT}$					$\lambda^2_{SEQ} >^{***} \lambda^1_{SEQ}$ $\lambda^1_{SEQ} <^{***} \lambda_{CHAT}$				

Note: The "<" and ">" signs summarize the results of statistical tests. ***, **, and * indicates significance at the 1%, 5%, and 10% level, respectively. A "<" and ">" sign without stars indicate no significant difference at the 10% level.

Table 2: Summary statistics and test results: Mutual cooperation rates

First repeated game

First round					All rounds				
SEQ		**SIM**		**CHAT**	**SEQ**		**SIM**		**CHAT**
27.54	$>^{***}$	8.11	$<^{***}$	61.76	31.34	$>^{***}$	11.76	$<^{***}$	64.56
27.54		$<^{***}$		61.76	31.34		$<^{***}$		64.56

All repeated games

First round					All rounds				
SEQ		**SIM**		**CHAT**	**SEQ**		**SIM**		**CHAT**
19.25	$>^{***}$	4.41	$<^{***}$	69.86	19.92	$>^{***}$	4.99	$<^{***}$	66.37
19.25		$<^{***}$		69.86	19.92		$<^{***}$		66.37

Note: The "<" and ">" signs summarize the results of statistical tests. *** indicate significance at the 1%

the numbers of messages sent and the content of the communication. We will then discuss in detail the relationship between the content of messages sent/received and the actions chosen.

Table 7 in the Online Appendix shows some statistics regarding the average number of messages sent per subject in each round of a match. The average number of messages sent in the first round of each match is 2.23. In later rounds, the average number of messages sent decreases but is always greater than 1.2.[17]

To find out about the content of messages subjects sent, we hired two external independent coders. Their first task was to carefully read the chats of each subject in each round of each match and assign one of several pre-selected verbal codes to each of them so that the assigned verbal code would best summarize the content of each individual chat. The coders had knowledge of the instructions of treatment CHAT, but at no time access to the actual data of this treatment.

Table 10 in the Online Appendix shows an overview of the content of chats as categorized by the two coders. Note that Table 10 only contains subjects' chats of rounds for which both coders assigned the same verbal code listed in the first column of this table. The rate of agreement was 75.71 percent of all individual chats (per subject, round and match). The most common chat content

[17] Half of the messages sent in later rounds such as rounds 6-8 consist of trivial exchange or the suggestion that both players choose cooperation again.

(32.04 percent of all cases considered in Table 10) was the suggestion that both players choose to cooperate. The second most common chat content (19.04 percent) was that a player agreed to a proposal made. In roughly two thirds of the cases in which a subject's chat was classified as "agree" the other player had made the suggestion that both choose to cooperate (verbal code "both1"). In 11.56 percent of the cases, a subject chose not to communicate in a round. In another 10.12 percent of the cases, a subject sent messages with trivial content. These four categories summarize 72.46 of all chats considered in Table 10.

As discussed in the Theory Section, the possibility to exchange messages via chat has the potential of greatly reducing strategic risk regarding the intended action of the other player *especially* in the first round of a match. According to (a nuanced interpretation of) our theory, a player is more likely to cooperate, the stronger the statement of her partner that she wants to (or *will*) cooperate. Likewise, a subject is more likely to cooperate, the stronger the statement of the subject herself that she wants to (or *will*) cooperate.

To see whether these basic mechanisms are borne out by the data, we assigned a second task to the external coders. We asked them to assess each round's chat of a subject according to a five-point scale, such that a higher code corresponds to a stronger signal for the subject's willingness to cooperate. More precisely, if a subject sent a weak or a strong signal indicating willingness to defect, this was coded as 1 for strong and 2 for weak. If a subject did not communicate a clear signal about intended play, this was coded as 3. Finally, if a subject sent a weak or a strong signal indicating willingness to cooperate, this was coded as 4 for weak and 5 for strong. A sixth coding option was when a subject's chat content was classified as saying "same" (see Table 10 in the Online Appendix). In this case, a coder would not know whether 1, 2, 4, or 5 was the appropriate code as the chosen actions were unknown to coders. In this case, coders were asked whether they could again say whether this was a weak or a strong suggestion to play the same action again. Using the action chosen in the previous round, this information was then used by us to assign 1, 2, 4, or 5 to those chats. Clearly, whether or not the content of a message could be categorized as a weak or strong signal indicating willingness to cooperate or defect is a subjective matter for the two coders.[18]

It turned out that for this task the coders only agreed in about 68 percent of the cases. To a substantial extent, the disagreement concerned those messages that were classified as either a

[18] The two coders had to code 5,180 individual chats (for all subjects, matches and rounds). While all of these chats were coded verbally (see above), we had to discard 75 data points due to mistakes made by the coders regarding the categorization just described.

weak or a strong signal indicating willingness to cooperate. In our analysis, we want to relate pairs of messages (*i.e.*, their codes) to the probability of cooperation and coordination on cooperation. However, the relatively high disagreement rate in the coding of individual chats would mean a substantial loss of data. We, therefore, decided to use a three-code system instead. We assigned code 1 to a chat that was originally coded as either 1 or 2 (weak or strong "defective" signal), code 2 to a chat that was coded as 3 ("neutral" signal), and code 3 to a chat that was originally coded as either 4 or 5 (weak or strong "cooperative" signal). In what follows we only use data for which the coders agree using this three-code system, which occurred in 90.24 percent of all valid individual assessments.

In Table 3, we show summary statistics of the relationship between pairs of messages exchanged and cooperation rates (top panels), mutual cooperation rates (middle panels), and the distribution of all mutual cooperation outcomes over pairs of messages exchanged (bottom panels). In this table, we distinguish between results in first rounds of all matches (left) and all rounds of all matches (right), as communication has a particularly important role in reducing strategic risk in the first round of a match but may not be essential after the first round.

We first comment on the summary statistics that relate pairs of signals exchanged to individual cooperation rates. These results are shown in the top (left and right) panels of Table 3. Note that these top panels are not symmetric, as cooperation rates might (and do) differ for "asymmetric" pairs of signals. Table 3 shows that the cooperation rate in the first round of each match was 0, 0.3, and 0.91 if *both* players sent defective, neutral, or cooperative messages, respectively. Hence, the mutual promise of cooperation boosts actual cooperation in first rounds dramatically. The first-round cooperation rate is pretty stable across matches if both players send cooperative signals. However, it decreases to 0 across matches if both players send neutral signals (starting from about 0.5 during the first 5 matches). So over time, first-round cooperation crucially relies on cooperative signals. Taking all rounds into account, we observe that the cooperation rate was 0.03, 0.53, and 0.92 if *both* players sent defective, neutral, or cooperative signals, respectively. Thus, the cooperation-boosting effect of both players sending cooperative signals appears to be maintained throughout a match. When all rounds of a match are considered, the average cooperation rate remains relatively high across matches even if both players sent neutral messages (in contrast to the evolution of cooperation in first rounds if both players sent neutral messages). To understand this pattern, note the following two observations which are also consistent with our theory. First, the share of "neutral" messages summarized as, e.g., "none" and "trivial" by the two coders (see

Table 3: Summary statistics: Cooperation and mutual cooperation rates

	Rounds 1 only					**All rounds**			
Cooperation rate depending on pairs of messages sent									
Own Intent	Other Intent				Own Intent	Other Intent			
↓	Defect	Neutral	Cooperate	Total	↓	Defect	Neutral	Cooperate	Total
Defect	0.00	0.15	0.06	0.07	Defect	0.03	0.09	0.13	0.09
	(12)	(13)	(18)	(43)		(30)	(47)	(39)	(116)
Neutral	0.08	0.30	0.43	0.36	Neutral	0.17	0.53	0.56	0.53
	(13)	(100)	(147)	(260)		(47)	(898)	(367)	(1,312)
Cooperate	0.33	0.53	0.91	0.88	Cooperate	0.49	0.63	0.92	0.88
	(18)	(147)	(1,768)	(1,933)		(39)	(367)	(2,773)	(3,179)
Total	0.16	0.42	0.87	0.80	Total	0.24	0.54	0.87	0.76
	(43)	(260)	(1,933)	(2,236)		(116)	(1,312)	(3,179)	(4,607)
Mutual cooperation rate depending on pairs of messages sent									
Own Intent	Other Intent				Own Intent	Other Intent			
↓	Defect	Neutral	Cooperate		↓	Defect	Neutral	Cooperate	
Defect	0.00				Defect	0.00			
	(12)					(30)			
Neutral	0.08	0.18			Neutral	0.02	0.47		
	(26)	(100)				(94)	(898)		
Cooperate	0.00	0.29	0.85		Cooperate	0.10	0.46	0.86	
	(36)	(294)	(1,768)			(78)	(734)	(2,773)	
Distribution of mutual cooperation outcomes over pairs of messages sent									
Own Intent	Other Intent				Own Intent	Other Intent			
↓	Defect	Neutral	Cooperate		↓	Defect	Neutral	Cooperate	
Defect	0				Defect	0			
	(0)					(0)			
Neutral	< 0.01	0.01			Neutral	< 0.01	0.14		
	(2)	(18)				(2)	(426)		
Cooperate	0	0.05	0.93		Cooperate	< 0.01	0.11	0.76	
	(0)	(86)	(1,496)			(8)	(334)	(2,377)	

Notes: Relationship between pairs of messages exchanged and the shares of cooperative choices (top), shares of mutual cooperation outcomes (middle), and distributions of all mutual cooperation outcomes (bottom). Results in first rounds of matches (left) and all rounds of matches (right). Numbers of observations in parentheses.

Table 10) increases substantially across rounds of a match. Second, the two regression analyses below will show that "actions speak louder than words," in the sense that once (mutual) cooperation has been observed, this is more predictive of (mutual) cooperation in the next round than promises of cooperation.[19] These two observations together suggest that the relatively high share of cooperation following the mutual exchange of neutral messages is driven by the early establishment of cooperation, which is maintained within matches even if only "neutral" messages are exchanged from then onwards. Finally, note that by holding the subject's own (partner's) message constant and then moving horizontally (vertically) in the top panels in Table 3, the cooperation rate increases monotonically with the message received (sent) with only one exception.

How does coordination on cooperation relate to observed pairs of messages? This is shown in the middle (left and right) panels in Table 3. Here we show the mutual cooperation rates for each possible pair of signals. Note that these panels are triangular matrices, as data stemming from asymmetric pairs of signals (e.g., cooperate-defect and defect-cooperate) are pooled. Focusing on first rounds of all matches only, the mutual cooperation rate was 0, 0.18, and 0.85 after *mutually* defective, neutral or cooperative signals, respectively (however, recall that the first-round cooperation rate, and thus, also the first-round mutual cooperation rate, falls to 0 with experience if neutral signals are exchanged). Considering all rounds of all matches, the mutual cooperation rate was 0, 0.47, and 0.86 after *mutually* defective, neutral or cooperative messages, respectively. The increase in the mutual cooperation rate given a pair of neutral messages when moving from first-round to all-rounds statistics is not surprising given the increased cooperation rate for this message pair when moving from first-round to all-rounds statistics, the reasons of which we discussed above.

Finally, we note that of all mutual cooperation outcomes observed in first rounds (all rounds), 93 percent (76 percent) were achieved after the exchange of cooperative messages. This information is shown in the bottom (left and right) panels of Table 3. When all rounds are considered, 14 percent of mutual cooperation outcomes were also the result of the mutual exchange of neutral messages. Again, this is driven by the early establishment of mutual cooperation, which makes exchanging cooperative signals less needed from then onwards.

To formally test for the results presented in Table 3, we will proceed by reporting the results of two sets of probit regressions. In both sets of regressions, we cluster all observations at the matching group level and report marginal effects.

[19] The crucial elements of our theory are: first-round cooperation happens only if cooperative messages are exchanged, and continuation of cooperation continues only if cooperation was chosen by both partners in the first round.

Table 4: Regressing actions on both players' communicated intents

	(1)	(2)	(3)	(4)
Own intent -	Rd = 1	Rd $\geq$ 1	Rd > 1	Rd > 1
Other intent	Match $\geq$ 1	Match $\geq$ 1	Match > 1	Match > 1
Defect-Defect	All-Defect	-0.661^{***}	-0.612^{***}	-0.711^{***}
		(0.078)	(0.097)	(0.068)
Defect-Neutral	-0.145	-0.530^{***}	-0.602^{***}	-0.608^{***}
	(0.117)	(0.067)	(0.111)	(0.138)
Defect-Cooperate	-0.363	-0.446^{***}	-0.436^{***}	-0.309^{*}
	(0.225)	(0.152)	(0.156)	(0.182)
Neutral-Defect	-0.297^{**}	-0.376^{**}	-0.360^{**}	-0.250
	(0.139)	(0.160)	(0.176)	(0.221)
Neutral-Cooperate	0.069^{*}	0.018	0.061	0.064
	(0.036)	(0.051)	(0.076)	(0.048)
Cooperate-Defect	0.021	-0.035	0.012	0.057
	(0.049)	(0.092)	(0.099)	(0.049)
Cooperate-Neutral	0.106^{***}	0.062	0.099	0.048
	(0.021)	(0.061)	(0.084)	(0.065)
Cooperate-Cooperate	0.605^{***}	0.390^{***}	0.363^{***}	0.275^{***}
	(0.021)	(0.114)	(0.134)	(0.081)
Other action in previous				0.647^{***}
round of current match				(0.058)
Observations	2,224	4,607	2,293	2,293

Notes: The dependent variable is Pr(cooperate). Marginal effects reported. The entry "All-Defect" indicates that for the corresponding pair of messages and slice of the data only defect choices were observed. ***, **, and * indicates significance at the 1%, 5%, and 10% level, respectively.

Table 4 presents the relationship between pairs of signals exchanged and the likelihood of a cooperative choice. For this purpose, we ran probit regressions with the dependent variable being whether or not cooperation was chosen in a round (coded as 1 or 0, respectively). The independent variables are binary variables coding the various pairwise combinations of signals. Note that in these signal pairs the order has significance, because we want to assess how players react to combinations of their partner's and own messages, consistent with the statistics presented in the top panels of Table 4. In our regression, we choose the behavior following the mutual exchange of neutral signals as the reference group and measure the effect of all independent variables with reference to this group. Note that if only first-round data of all matches are considered, subject behavior for the defect-defect signal pair was constant (*i.e.*, all-defect choices). This implies that these observations are dropped from the probit regressions. We indicate this in Table 4 by the entry "All-Defect." To save space, we focus on the most important features in Table 4.

In column (1) of Table 4, we measure the relationship between signal pairs and the likelihood of cooperation in the first round of each match. Consistent with the numbers shown in Table 3, signal pairs that involve at least one defective signal do not foster cooperation. However, mutual exchange of cooperative signals significantly improves cooperation relative to the reference group of neutral-neutral message pairs.[20] In column (2) of Table 4 we include all rounds and observe, for instance, that various signal pairs involving defective signals significantly reduce cooperative choices relative to the reference group. In column (4) we drop all data stemming from first rounds of a match. As an additional regressor, we include current partner's choice in the previous round of the current match. (For a clean comparison, column (3) reports the results for the same selection of data without this additional regressor.) We observe that if the partner cooperated in the previous round, this boosts the likelihood of cooperation by about 65 percent. This coefficient is also significantly larger (pairwise Wald tests) than any other estimated coefficient in column (4), which suggests the interpretation that "actions speak louder than words" after the first round consistent with our theory. We also observe that "own-defect" signal pairs significantly reduce cooperative choices, while the exchange of cooperative signals significantly improves cooperation relative to the reference group.

We now analyze the relationship between signal pairs and the likelihood of coordination on cooperation. For this purpose, we ran probit regressions with the dependent variable being whether

[20] The estimated coefficient of "Cooperate-Cooperate" in column (1) of Table 4 is highly significantly larger than any other estimated coefficient in column (1) (pairwise Wald tests).

Table 5: Regressing mutual cooperation outcome both players' communicated intents

	(1) Rd = 1 Match ≥ 1	(2) Rd ≥ 1 Match ≥ 1	(3) Rd > 1 Match ≥ 1	(4) Rd > 1 Match ≥ 1
Defect-Defect	All-Zero	All-Zero	All-Zero	All-Zero
Neutral-Defect	−0.183	−0.638***	All-Zero	All-Zero
	(0.154)	(0.084)		
Cooperate-Defect	All-Zero	−0.452***	−0.347*	−0.125
		(0.137)	(0.185)	(0.145)
Cooperate-Neutral	0.101	−0.018	0.045	−0.011
	(0.070)	(0.065)	(0.081)	(0.046)
Cooperate-Cooperate	0.665***	0.394***	0.363***	0.210***
	(0.061)	(0.098)	(0.129)	(0.081)
Coordination in previous round of current match				0.758***
				(0.053)
Observations	1,095	2,289	1,142	1,142

Notes: The dependent variable is Pr(mutual cooperation outcome). Marginal effects reported. Reference group: "Neutral- Neutral." The entry "All-Zero" indicates that for the corresponding pair of messages and slice of the data mutual cooperation was never achieved. ***, **, and * indicates significance at the 1%, 5%, and 10% level, respectively.

or not a pair of subjects coordinated on cooperation (coded as 1 or 0, respectively). The independent variables are again binary variables coding the various pairwise combinations of signals. The results are presented in Table 5. Note that here the order of signal pairs listed in the first column of Table 5 are inconsequential, as in the middle panels of Table 3. We again choose the behavior following the exchange of a pair of neutral messages as the reference group and measure the effect of all independent variables with reference to this group. For some slices of the data and signal pairs, mutual cooperation was never achieved. This implies again that these observations are dropped from the probit regressions. We indicate this in Table 5 by the entry "All-Zero." Again, to save space we focus on the most important results of the regressions shown in Table 5.

In column (1) of Table 5, we consider the first round of each match. The main finding is that first-round coordination on cooperation is significantly more likely relative to the reference group *only if* a pair of players exchanged cooperative messages. This holds true (albeit to a less strong extent) when, in column (2), we include all rounds in the data. In column (4) we include a variable that indicates whether a pair of players managed to achieve mutual cooperation in the previous round of the current match. (For a clean comparison, column (3) reports the results for the same selection of data without this additional regressor.) The estimated effect of this new variable is at 0.758 "big" and highly significantly larger than both 0 and the estimated coefficient of the "Cooperate-Cooperate" signal pair (Wald test). This, again, suggests the interpretation that "actions speak louder than words." Once mutual cooperation was successfully achieved in the previous round of a match, this is more predictive of mutual cooperation in the current round than mutual promises of cooperation.

5.3 Strategies used

We use the structural frequency estimation method (SFEM) developed in DB&F to estimate the shares of several strategies used in the three treatments. To use this method, a set of repeated-game strategies needs to be specified and the incidence of each strategy is estimated via maximum likelihood. DB&F assume that subjects possibly make errors in executing a strategy; a parameter γ captures the amount of noise in the data, with choices becoming purely random as γ approaches infinity. We assume that subjects use five of the six strategies suggested in DB&F: Always Defect (AD), Always Cooperate (AC), Grim (G), Tit for Tat (TFT), Win Stay Loose Shift (WSLS) (start cooperating, and cooperate if both or neither player cooperated in the previous round, otherwise defect).[21] We disregard a sixth (trigger) strategy (called T2) as it works with a two-period memory, which is unlikely to play a role in our experiment as the average duration of a repeated game is just 2. Moreover, its share was estimated to be 0 in the corresponding treatment of DB&F. The five strategies were assumed to be used by players in SIM, CHAT, and SEQ (first movers only). For second movers in SEQ, strategies G, TFT and WSLS were adjusted: Strategies G and WSLS condition on the first-mover's current-period and the second mover's previous period choice; TFT conditions on the first mover's current-period choice. Table 6 shows the results using all data of the experiment. Note that the coefficient for WSLS is implied by the requirement that the proportions of strategies included must sum to one. As will become clear, most estimation results are in line with the theory in Section 3.

Let us start by comparing the estimated shares of strategies in SIM with those in the corresponding treatment in DB&F. The latter only find positive shares for strategies AD (0.92) and TFT (0.08). Our estimates for these strategies are very similar with AD at 0.902 and TFT

[21] Recall that our theory in Section 3 is based on cooperative strategies ("CS") that include G and TFT.

Table 6: Shares of estimated strategies (All data)

	SEQ		**SIM**	**CHAT**
	1st Mover	2nd Mover		
Always Defect	0.720***	0.379***	0.902***	0.064
	(0.116)	(0.108)	(0.106)	(0.054)
Always Cooperate	0.020	0.000	0.000	0.117
	(0.038)	(0.0000)	(0.000)	(0.138)
Grim Trigger	0.000	0.078	0.043	0.300**
	(0.080)	(0.073)	(0.072)	(0.126)
Tit-for-Tat	0.217*	0.543***	0.055	0.492***
	(0.127)	(0.115)	(0.067)	(0.144)
Win-Stay-Lose-Shift	0.043	0.000	0.000	0.031
γ	0.555***	0.385***	0.537***	0.552***
	(0.088)	(0.044)	(0.115)	(0.091)

Note: Data from all matches and rounds. ***, **, and * indicates significance at the 1%, 5%, and 10% level, respectively.

at 0.055. Additionally, we find the share of strategy G to be positive at 0.043. Note that the shares of strategy TFT in DB&F and G and TFT in our data are not significantly different from 0. We conclude that both in DB&F and in our treatment SIM subjects predominantly use the always-defect strategy.

Moving from SIM to CHAT, we see drastic changes in estimated shares of strategies. First, the share of AD decreases drastically from 0.902 in SIM to 0.064 in CHAT. Second, the shares of the cooperative strategies AC, G, and TFT, respectively, increase from 0, 0.043, and 0.055 in SIM to 0.117, 0.300, and 0.492 in CHAT. However, in CHAT only the shares of strategies G and TFT are statistically different from 0. Hence, the substantial increase of cooperative choices in CHAT in comparison to SIM is driven by the drastic drop in the always-defect and the significant increase in the (conditionally) cooperative strategies G and TFT.

Moving from SIM to first movers in SEQ, we note that the estimated share of strategy AD decreases from 0.902 in SIM to 0.702 in SEQ and the estimated share of strategy TFT increases from 0.055 in SIM to 0.217 in SEQ. Both changes are not significantly different. However, while the share of TFT in SIM is not significantly different from 0, the share of TFT for first movers in SEQ is (albeit only at the 10 percent level). These observations are consistent with the higher share of cooperative choices by first movers in SEQ compared with those in SIM. Finally, comparing the estimated shares of strategies by first and second movers in SEQ, we observe that the share of AD decreases from 0.720 for first movers to 0.379 for second movers. Moreover, the share of strategy TFT increases from 0.217 for first movers to 0.543 for second movers. Based on the z-statistic defined in (4) in the Appendix, both changes are significant at the 5 percent level. Hence, compared to first movers, second movers make less use of strategy AD and more use of strategy TFT.

6 Summary

The theory of infinitely repeated games is so far largely based on the assumption that players are motivated solely by their material self interest. Although this theory correctly predicts an increased tendency of players to cooperate in specific infinitely repeated games, the multiplicity of predicted equilibrium outcomes has been the subject of criticism by various authors (see, *e.g.*, Fudenberg and Maskin (1993) or Dal Bó and Fréchette (2011)). An additional shortcoming of this theory is that it cannot predict changes in behavior when specific aspects of the game change, such as the timing of players' moves or the possibility of players communicating with each other. In this paper we propose to mitigate such shortcomings by incorporating a privately observed, heterogeneous taste for cooperation into the context of infinitely repeated (dilemma) games. Our model captures an important component of strategic risk especially in those games in which cooperation is not an equilibrium or a risk-dominant action. Additionally, it allows us to make intuitive comparative static predictions regarding changes in (a) payoff parameters of a game, (b) the distribution of players' preferences, (c) the timing of players' moves, or (d) the mode of communication. While standard theory is unresponsive to such changes, we show that our model takes these changes into account. In particular, we show that variations of a standard simultaneous-move stage game reduce strategic risk by making it easier for players to recognize a cooperative opponent. Moreover, we provide new experimental evidence that is in line with these theoretical predictions. In future research we plan to further demonstrate the usefulness of our theory by applying it to other infinitely repeated games and variants thereof.

References

[1] Ahn, T.K., M.Lee, L. Ruttan, and J.Walker (2007): Asymmetric Payoffs in Simultaneous and Sequential Prisoner's Dilemma Games, *Public Choice* 132, 353–366.

[2] Andersson, O., C. Argenton, and J.W. Weibull (2014): Robustness to Strategic Uncertainty, *Games and Economic Behavior* 85, 272–288.

[3] Andreoni, J. and J.H. Miller (1993): Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence, *Economic Journal* 103, 570–85.

[4] Andreoni, J. and L. Samuelson (2006): Building Rational Cooperation, *Journal of Economic Theory* 127, 117–154.

[5] Aoyagi, M. and G.R. Fréchette (2009): Collusion as Public Monitoring Becomes Noisy: Experimental Evidence, *Journal of Economic Theory* 144, 1135–1165.

[6] Arechar, A., A. Dreber, D. Fudenberg, and D.G. Rand (2017): I'm just a soul whose intentions are good: The Role of Communication in Noisy Repeated Games, *Games and Economic Behavior* 124, 726–743.

[7] Barut, Y., D. Kovenock, and C. Noussair. (2002): Comparison of Multiple-Unit All-Pay and Winner-Pay Auctions Under Incomplete Information, *International Economic Review* 43, 675–707.

[8] Blonski, M., P. Ockenfels and G. Spagnolo (2011): Equilibrium Selection in the Repeated Prisoner's Dilemma: Axiomatic Approach and Experimental Evidence, *American Economic Journal: Microeconomics* 3, 164–192.

[9] Bolton, G.E. and A. Ockenfels (2000): ERC – A Theory of Equity, Reciprocity and Competition, *American Economic Review* 100, 166–193.

[10] Cabrales A., R. Miniaci, M. Piovesan, and G. Ponti (2010): Social Preferences and Strategic Uncertainty: An Experiment on Markets and Contracts, *American Economic Review* 100, 2261–2278.

[11] Camera, G. and M. Casari (2009) Cooperation Among Strangers Under the Shadow of the Future, *American Economic Review* 99, 979–1005.

[12] Camerer, C. and K. Weigelt (1988): Experimental Tests of a Sequential Equilibrium Reputation Model, *Econometrica* 56, 1–36.

[13] Charness, G. and M. Rabin (2002): Understanding Social Preferences with Simple Tests, *Quarterly Journal of Economics* 117, 817–869.

[14] Cho, I.K. and D.M. Kreps (1987): Signaling Games and Stable Equilibria, *Quarterly Journal of Economics* 102, 179–221

[15] Dal Bó, P. (2005): Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games, *American Economic Review* 95, 1591–1604.

[16] Dal Bó, P., and G.R. Fréchette (2011): The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence, *American Economic Review* 101, 411–429.

[17] Duffy, J. and J. Ochs (2009): Cooperative Behavior and the Frequency of Social Interaction, *Games and Economic Behavior* 66, 785–812.

[18] Dvořák, F. and S. Fehrler (2018): Negotiating Cooperation under Uncertainty: Communication in Noisy, Indefinitely Repeated Interactions, Working Paper, University of Konstanz.

[19] Engelmann, D. and H.-T. Normann (2010): Maximum Effort in the Minimum-Effort Game, *Experimental Economics* 13, 249–259.

[20] Engle-Warnick, J. and R. Slonim (2004): The Evolution of Strategies in a Trust Game, *Journal of Economic Behavior and Organization* 55, 553–573.

[21] Engle-Warnick, J. and R. Slonim (2006a): Learning to Trust in Indefinitely Repeated Games, *Games and Economic Behavior* 54, 95–114.

[22] Engle-Warnick, J. and R. Slonim (2006b): Inferring Repeated Game Strategies from Actions: Evidence from Trust Game Experiments, *Economic Theory* 28, 603–632.

[23] Embrey, M., G. R. Fréchette, and E. Stacchetti (2013): An Experimental Study of Imperfect Public Monitoring: Efficiency versus Renegotiation-Proofness, Working Paper, New York University.

[24] Fehr, E. and K.M. Schmidt (1999): A Theory of Fairness, Competition, and Cooperation, *Quarterly Journal of Economics* 117, 817–68.

[25] Fehr, E., M. Naef and K.M. Schmidt (2006): Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment, *American Economic Review* 96, 1912–1917.

[26] Fischbacher, U. (2007): Z-tree - Zurich toolbox for Readymade Economic Experiments, *Experimental Economics* 10, 171-178.

[27] Fréchette, G. R. and S. Yuksel (2017): Infinitely Repeated Games in the Laboratory: Four Perspectives on Discounting and Random Termination, *Experimental Economics* 20, 279–308.

[28] Fudenberg, D. and E. Maskin (1993): Evolution and Repeated Games, Unpublished Working Paper.

[29] Fudenberg, D., D.G. Rand, and A. Dreber (2012): Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World, *American Economic Review* 102, 720–749.

[30] Ghidoni, R. and S. Suetens (2018): Sequentiality Increases Cooperation in Repeated Prisoner's Dilemma, Tilburg University, mimeo.

[31] Greiner, B. (2015): Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE, *Journal of the Economic Science Association* 1, 114–125.

[32] Hayashi, N., E. Ostrom, J. Walker, and T. Yamagishi (1999): Reciprocity, trust, and the sense of control: A Cross-societal Study, *Rationality and Society* 11, 27–46.

[33] Heinemann, F., R. Nagel, and P. Ockenfels (2009): Measuring Strategic Uncertainty in Coordination Games, *Review of Economic Studies* 76, 181–221.

[34] Kartal, M., W. Müller, and J. Tremewan (2017): Building Trust: The Costs and Benefits of Gradualism, Working Paper, University of Vienna.

[35] Kartal, M. (2018): Honest Equilibria in Reputation Games: The Role of Time Preferences, *American Economic Journal: Microeconomics* 10, 278-314.

[36] Kreps, D., P. Milgrom, J. Roberts, and R. Wilson (1982): Rational Cooperation in the Finitely Repeated Prisoner's Dilemma, *Journal of Economic Theory* 27, 245–252.

[37] Khadjavi, M. and A. Lange (2013): Prisoners and their Dilemma, *Journal of Economic Behavior and Organization* 92, 163–175.

[38] Levine, D.K. (1998): Modeling Altruism and Spitefulness in Experiments, *Review of Economic Dynamics* 1, 593–622.

[39] Morris,S. and H.S. Shin (2002): Measuring Strategic Uncertainty, mimeo.

[40] Mermer, A.G., W. Müller, and S. Suetens (2018): Cooperation in Indefinitely Repeated Games of Strategic Complements and Substitutes, Working Paper, University of Vienna.

[41] Noussair, C., C. Plott, and R. Riezman (1995): An Experimental Investigation of the Patterns of International Trade, *American Economic Review* 85, 462–91.

[42] Oskamp, S. (1974): Comparison of Sequential and Simultaneous Responding, Matrix, and Strategy Variables in a Prisoner's Dilemma Game, *Journal of Conflict Resolution* 18, 107–116.

[43] Palfrey, T.R. and H. Rosenthal (1994): Repeated Play, Cooperation, and Coordination: An Experimental Study, *Review of Economic Studies* 61, 545–65.

[44] Rabin, M. (1993): Incorporating Fairness into Game Theory and Economics, *American Economic Review* 83, 1281–1302.

[45] Selten, R. (1975): Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games, *International Journal of Game Theory* 4, 25–55.

[46] Vespa, E. (2015): An Experimental Investigation of Strategies in the Dynamic Common Pool Game, Working Paper, Available at SSRN: https://ssrn.com/abstract=1961450.

[47] Vespa, E. and A.J. Wilson (2017): Information Transmission under the Shadow of the Future: An Experiment, Working Paper University of California at Santa Barbara.

[48] Van Huyck, John B., Raymond C. Battalio, and Richard O. Beil (1990): Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure, *American Economic Review* 80, 234–48.

[49] Van Huyck, John B., Raymond C. Battalio, and Richard O. Beil (1991): Strategic Uncertainty, Equilibrium Selection, and Coordination Failure in Average Opinion Games, *Quarterly Journal of Economics* 106, 885–910.

A Omitted Theory Materials

We first present and prove the following claim, which we repeatedly invoke in the proofs of our propositions below.

Claim 1 *If* $\Pi(\delta, p, \gamma) \geq 0$ *then* $\Pi(\delta, p', \gamma) > 0$ *for all* $p' > p$ *and if* $\Pi(\delta, p, \gamma) \leq 0$*, then* $\Pi(\delta, p', \gamma) < 0$ *for all* $p' < p$*.*

Proof: $\Pi(\delta, p, \gamma)$ can be written as

$$p\left(\frac{u(C,C,\gamma)}{1-\delta} - u(D,C,\gamma) - \delta\frac{u(D,D,\gamma)}{1-\delta}\right) + (1-p)(u(C,D,\gamma) - u(D,D,\gamma)).$$

If $\Pi(\delta, p, \gamma) \geq 0$, then the first term inside parentheses must be strictly positive (because the term $u(C,D,\gamma) - u(D,D,\gamma)$ is negative); *i.e.*,

$$\frac{u(C,C,\gamma)}{1-\delta} > u(D,C,\gamma) + \delta\frac{u(D,D,\gamma)}{1-\delta}.$$

Hence, increasing p strictly increases $\Pi(\delta, p, \gamma)$. If $\Pi(\delta, p, \gamma) \leq 0$, then $\Pi(\delta, p', \gamma) < 0$ for $p' < p$ because a decrease in p increases the weight of the term $u(C,D,\gamma) - u(D,D,\gamma)$, which is negative. Thus, $\Pi(\delta, p', \gamma)$ cannot be nonnegative if $\Pi(\delta, p, \gamma) \leq 0$ and $p' < p$.

Proof of Proposition 1 (i). First note that equilibrium must have a cutoff form by A2. Next, we show that there always exists a $\gamma^* \in (\underline{\gamma}, \bar{\gamma})$ such that $\Pi(\delta, 1-F(\gamma^*), \gamma^*) = 0$. By A1S, there exist γ_1 and γ_2 such that $\gamma_2 < \gamma_1$, $\Pi(\delta, 1-F(\gamma_1), \gamma_1) > 0$ and $\Pi(\delta, 1-F(\gamma_2), \gamma_2) < 0$. The latter inequality holds because $\Pi(\delta, 1, \gamma_2) < 0$ implies that $\Pi(\delta, 1-F(\gamma_2), \gamma_2) < 0$ by Claim 1. Thus, by continuity of $\Pi(\delta, 1-F(\gamma), \gamma)$ in γ, there must exist a $\gamma^* \in (\underline{\gamma}, \bar{\gamma})$ such that $\Pi(\delta, 1-F(\gamma^*), \gamma^*) = 0$. Next, we show that an asymmetric equilibrium is not possible. Suppose towards a contradiction that there exist asymmetric equilibrium cutoffs γ_1^* and γ_2^* for Players 1 and 2, and assume without loss of generality that $\gamma_1^* > \gamma_2^*$. Then, in equilibrium we must have $\Pi(\delta, 1-F(\gamma_2^*), \gamma_1^*) \leq 0$, and $\Pi(\delta, 1-F(\gamma_1^*), \gamma_2^*) \geq 0$ (we allow for the possibility that $\gamma_1^* = \bar{\gamma}$ or $\gamma_2^* = \underline{\gamma}$). From $\Pi(\delta, 1-F(\gamma_2^*), \gamma_1^*) \leq 0$, $\gamma_1^* > \gamma_2^*$, and A2, it follows that

$$\Pi(\delta, 1-F(\gamma_2^*), \gamma_2^*) < 0. \tag{1}$$

Since $\Pi(\delta, 1-F(\gamma_1^*), \gamma_2^*) \geq 0$ and $\Pi(\delta, p, \gamma_2^*)$ is continuous in p, there must exist a $\gamma \in (\gamma_2^*, \gamma_1^*]$

such that $\Pi(\delta, 1 - F(\gamma), \gamma_2^*) = 0$. However, by Claim 1, $\Pi(\delta, 1 - F(\gamma), \gamma_2^*) = 0$ implies that $\Pi(\delta, 1 - F(\gamma_2^*), \gamma_2^*) > 0$ as $\gamma > \gamma_2^*$ and $1 - F(\gamma_2^*) > 1 - F(\gamma)$, a contradiction to (1) above. Finally, it can easily be checked that no type $\gamma > \gamma^*$ has an incentive to deviate from CS to AD, and likewise no type $\gamma < \gamma^*$ benefits from deviating from AD, for example choosing C in the first period (or in the first few periods) and then reverting to AD.

Proof of Proposition 1 (ii). Let $\gamma(a)$ denote the minimum equilibrium cutoff in game $\Gamma(a, b, c, d, \delta)$ and assume that $a' < a$. We will show that any equilibrium cutoff in game $\Gamma(a', b, c, d, \delta)$ must be strictly higher than $\gamma(a)$.[22] Let $\Pi_a(\delta, p, \gamma)$ and $\Pi_{a'}(\delta, p, \gamma)$ denote $\Pi(\delta, p, \gamma)$ in games $\Gamma(a, b, c, d, \delta)$ and $\Gamma(a', b, c, d, \delta)$, respectively. Suppose towards a contradiction that there exists an equilibrium cutoff γ' in game $\Gamma(a', b, c, d, \delta)$ such that $\gamma' \leq \gamma(a)$. As $a > a'$ and $\Pi_{a'}(\delta, 1 - F(\gamma'), \gamma') = 0$, it follows that $\Pi_a(\delta, 1 - F(\gamma'), \gamma') > 0$. Moreover, there must exist a type $\hat{\gamma} < \gamma'$ such that $\Pi_a(\delta, 1 - F(\hat{\gamma}), \hat{\gamma}) < 0$. To see why, note that by A1S there exists a type $\gamma_2 > \underline{\gamma}$ such that $\Pi_a(\delta, 1, \gamma_2) < 0$. By A2, $\Pi_a(\delta, 1, \gamma) < 0$ must hold for every $\gamma < \gamma_2$. By Claim 1, $\Pi_a(\delta, 1, \gamma) < 0$ implies that $\Pi_a(\delta, 1 - F(\gamma), \gamma) < 0$. In particular, $\Pi_a(\delta, 1 - F(\gamma), \gamma) < 0$ holds for every $\gamma < \gamma_2$. As a result, there must exist a type $\hat{\gamma} < \gamma'$ such that $\Pi_a(\delta, 1 - F(\hat{\gamma}), \hat{\gamma}) < 0$. Then by continuity, there exists a $\gamma \in (\hat{\gamma}, \gamma')$ such that $\Pi_a(\delta, 1 - F(\gamma), \gamma) = 0$ and $\gamma < \gamma(a)$, in contradiction to $\gamma(a)$ being the minimum equilibrium cutoff in $\Gamma(a, b, c, d, \delta)$. The proof for the case in which $b < b'$ is very similar, and therefore omitted.

Proof of Proposition 1 (iii). The proof is very similar to the proof of part (ii), and therefore omitted.

Proof of Proposition 1 (iv). Let γ^* denote the minimum equilibrium cutoff in game $\Gamma(a, b, c, d, \delta)$ with preferences characterized by $F(\gamma)$. We now show that the minimum equilibrium cutoff induced by $F'(\gamma)$ is strictly lower than γ^*. Since $1 - F(\gamma^*) < 1 - F'(\gamma^*)$ (by first order stochastic dominance) and $\Pi(\delta, 1 - F(\gamma^*), \gamma^*) = 0$, it follows from Claim 1 that $\Pi(\delta, 1 - F'(\gamma^*), \gamma^*) > 0$. By A1S and A2, there exists a type $\gamma_2 > \underline{\gamma}$ such that $\Pi(\delta, 1, \gamma) < 0$ for every $\gamma < \gamma_2$. In particular, $\Pi(\delta, 1 - F'(\gamma), \gamma) < 0$ for every $\gamma < \gamma_2$ by Claim 1. Hence by continuity there exists $\hat{\gamma} \in (\gamma_2, \gamma^*)$ such that $\Pi(\delta, 1 - F'(\hat{\gamma}), \hat{\gamma}) = 0$. Since $\hat{\gamma} < \gamma^*$, the desired result follows.

Proof of Proposition 2. Let γ^* denote the minimum equilibrium cutoff type in the infinitely repeated *simultaneous* prisoners' dilemma game. Let γ_1^* and γ_2^* denote the repeated sequential

[22] Let $\gamma(a) = \inf\{\gamma | \Pi_a(\delta, p, \gamma) = 0\}$. By continuity, $\Pi_a(\delta, p, \gamma(a)) = 0$ must hold. Thus, $\gamma(a) = \min\{\gamma | \Pi_a(\delta, p, \gamma) = 0\}$.

prisoner's dilemma game equilibrium cutoff for the first mover and the second mover respectively. We first consider the case in which $\gamma_1^* < \bar{\gamma}$ (see our argument regarding uniqueness below). Note that in that case $\gamma_2^* < \bar{\gamma}$ and must satisfy $\Pi(\delta, 1, \gamma_2^*) = 0$ because A1S and Claim 1 hold, and because, in equilibrium, a first-mover choice of C implies a commitment to CS, which is optimal; *i.e.*, choosing C and switching to D in a later period although the second mover always reciprocates cooperation by the first mover is strictly dominated for all first movers for whom $\Pi(\delta, 1 - F(\gamma_2^*), \gamma) \neq 0$. In particular, it is dominated by CS for those for whom $\Pi(\delta, 1 - F(\gamma_2^*), \gamma) > 0$ and by AD for those for whom $\Pi(\delta, 1 - F(\gamma_2^*), \gamma) < 0$. A similar argument is also true for second movers; choosing C and then switching to D later against a cooperating first-mover is dominated. Thus, in line with our simplifying assumption, players optimally select into one of two strategies (CS or AD) according to their type. To prove (i), we first show that $\gamma_2^* < \gamma^*$, which is part of the statement in (ii). Assume towards a contradiction that $\gamma_2^* \geq \gamma^*$. Since $\Pi(\delta, 1, \gamma_2^*) = 0$, it follows that $\Pi(\delta, 1, \gamma^*) \leq 0$ by A2. However, this implies that $\Pi(\delta, p, \gamma^*) < 0$ for all $p < 1$. In particular, $\Pi(\delta, 1 - F(\gamma^*), \gamma^*) < 0$, a contradiction. Hence, we have shown that $\gamma_2^* < \gamma^*$. Next, we show that $\gamma_1^* < \gamma^*$. Suppose towards a contradiction that $\gamma_1^* \geq \gamma^*$. Note that γ_1^* must be interior by A1S and Claim 1; *i.e.,* $\Pi(\delta, 1 - F(\gamma_2^*), \gamma_1^*) = 0$. By A2, it follows that $\Pi(\delta, 1 - F(\gamma_2^*), \gamma^*) \leq 0$. But then $\Pi(\delta, p, \gamma^*) < 0$ for all $p < 1 - F(\gamma_2^*)$ by Claim 1. In particular, $\Pi(\delta, 1 - F(\gamma^*), \gamma^*) < 0$ since $1 - F(\gamma^*) < 1 - F(\gamma_2^*)$, which is a contradiction. Hence, $\gamma_1^* < \gamma^*$ holds as well. As a result, the first-mover cooperation rate in the first period as well as the continuation game is strictly higher in the sequential game than the corresponding cooperation rates in the simultaneous game, and part (i) is proved. To prove part (ii), we will show, in addition to what we have already shown, that $\gamma_2^* < \gamma_1^*$. Suppose towards a contradiction that $\gamma_2^* \geq \gamma_1^*$. Since we must have $\Pi(\delta, 1, \gamma_2^*) = 0$, and $\Pi(\delta, 1 - F(\gamma_2^*), \gamma_1^*) = 0$ in equilibrium, and $\gamma_2^* \geq \gamma_1^*$ by hypothesis, it follows that $\Pi(\delta, 1 - F(\gamma_2^*), \gamma_2^*) \geq 0$ by A2 and $\Pi(\delta, 1, \gamma_2^*) > 0$ by Claim 1, which is a contradiction. Finally, the proof of part (iii) follows from the fact that $(1 - F(\gamma_1^*))(1 - F(\gamma_2^*)) > (1 - F(\gamma^*))^2$ as $\gamma_1^* < \gamma^*$ and $\gamma_2^* < \gamma^*$. Hence the proposition is proved.

Assuming that $\gamma_1^* < \bar{\gamma}$, we have a unique equilibrium. To see why, recall that $\Pi(\delta, 1, \gamma_2^*) = 0$ must hold for the second mover (and the second mover's expectation that the first mover will continue cooperating is correct) since only those first movers who find CS better than AD will choose C in the first period as argued above. As a result, for those first-mover types who prefer CS over AD, $\Pi(\delta, 1 - F(\gamma_2^*), \gamma_1^*) = 0$ must hold. But since γ_2^* is unique, γ_1^* is also unique. The other possibility is the case in which $\gamma_1^* = \bar{\gamma}$. If no cooperation outcome is an equilibrium, then the first mover choice of

C in the first round is an off-the-equilibrium path move and is not reciprocated because the second mover assumes that C is a mistake and will not be chosen again. The case in which $\gamma_1^* = \bar{\gamma}$ cannot be an equilibrium if there exists a sufficiently cooperative type γ such that $\Pi(\delta, \mu, \gamma) = 0$, where μ denotes the measure of γ-types such that $u(C, C, \gamma) > u(D, C, \gamma)$ holds. Even if there is no such type, we still argue that a no-cooperation equilibrium is "unreasonable" in the spirit of the Intuitive Criterion of Cho and Kreps (1987). The sketch of the proof of this claim proceeds by integrating the expectations of first movers and second movers regarding their partners actions' into the type space. In a reasonable equilibrium, a first mover who chooses C must be assigned the belief that the first mover is highly cooperative *and* sufficiently optimistic regarding the chances that the second mover will reciprocate. This is because only a first mover with such preferences and optimistic beliefs benefits from choosing C in the first period relative to the outcome in the equilibrium with $\gamma_1^* = \bar{\gamma}$ along the lines of a dynamic version of the Intuitive Criterion (see Kartal (2018) for a formal dynamic application of the Intuitive Criterion). In a similar vein, a first mover who chooses C and is reciprocated must believe that the second mover is highly cooperative and sufficiently optimistic that the first mover will continue cooperating because otherwise reciprocating the first mover's cooperation is dominated by defection. Given this construction, it is a best response for a cooperative second mover to be optimistic and reciprocate the first mover's cooperation choice in the first period, and thus, a highly cooperative first mover must deviate and choose to cooperate in the first round. Hence, under assumption A1S, an equilibrium in which there is no cooperation in the sequential game is unreasonable.

Proof of Proposition 3. We start with the benchmark case in which lying is costly enough so that no type misrepresents her incentive to cooperate. Let γ^* be such that $\Pi(\delta, 1, \gamma^*) = 0$ holds. We construct an equilibrium with a cutoff of γ^* as follows. Every type $\gamma > \gamma^*$ communicates either suggesting mutual cooperation or agreeing after the other party suggests it (depending on who starts the communication). After a type-γ player with $\gamma > \gamma^*$ suggests mutual cooperation, type-γ player indeed cooperates if the matched player agrees (so there is mutual agreement on cooperation), and chooses to defect otherwise. If the matched player suggests mutual cooperation, then every type $\gamma > \gamma^*$ agrees and cooperates, whereas every other type disagrees. No type $\gamma < \gamma^*$ suggests (mutual) cooperation and defects afterwards as lying is costly. It is easy to show that there is no incentive to deviate from this for any type. No type $\gamma > \gamma^*$ has an incentive to defect and deviate from the mutual agreement on cooperation since $\Pi(\delta, 1, \gamma) > 0$ and lying is costly. No type $\gamma < \gamma^*$ has an incentive to misrepresent her intentions and defect since lying is sufficiently costly.

Next, we consider the case in which a positive fraction of types is dishonest. Throughout, we will assume that every type who will or intends to choose the cooperative strategy communicates her intention truthfully. Thus, dishonesty refers to an intention to choose AD but pretending to be a cooperative type in the communication. Let $\rho(\gamma')$ denote the fraction of dishonest γ types such that $\gamma < \gamma'$. Note that by Bayesian updating, a type above the equilibrium cutoff γ^* will know that upon mutual agreement on cooperation the other player will cooperate with probability $\frac{1-F(\gamma^*)}{1-F(\gamma^*)+F(\gamma^*)\rho(\gamma^*)}$. This is because $\rho(\gamma^*)$ fraction of types below the cutoff will imitate a cooperator and pretend to agree on mutual cooperation. Thus, the cutoff in the most cooperative equilibrium of the communication games is such that $\Pi(\delta, \frac{1-F(\gamma^*)}{1-F(\gamma^*)+F(\gamma^*)\rho(\gamma^*)}, \gamma^*) = 0$. The full honesty scenario boils down to the case in which $\rho(\gamma^*) = 0$ and $\Pi(\delta, 1, \gamma^*) = 0$ holds. At the other extreme, we have $\rho(\gamma^*) = 1$ and $\Pi(\delta, 1 - F(\gamma^*), \gamma^*) = 0$. In this case, communication is entirely uninformative (babbling), and the most cooperative equilibrium is identical in the games with and without communication. For any $\rho(\gamma) < 1$, the equilibrium cutoff with communication is strictly lower than that in the simultaneous game. To show this, let $\hat{\gamma}$ denote the minimum equilibrium cutoff in the simultaneous game. Since $\Pi(\delta, 1 - F(\hat{\gamma}), \hat{\gamma}) = 0$ and $\rho(\hat{\gamma}) < 1$, it follows that $\Pi(\delta, \frac{1-F(\hat{\gamma})}{1-F(\hat{\gamma})+F(\hat{\gamma})\rho(\hat{\gamma})}, \hat{\gamma}) > 0$ from Claim 1. Moreover, by A1S and A2, there exists a type $\gamma_2 > \underline{\gamma}$ such that $\Pi(\delta, 1, \gamma) < 0$ for every $\gamma < \gamma_2$. In particular, $\Pi(\delta, \frac{1-F(\gamma)}{1-F(\gamma)+F(\gamma)\rho(\gamma)}, \gamma) < 0$ for every $\gamma < \gamma_2$ by Claim 1. Thus, $\gamma_2 < \hat{\gamma}$ and there exists a $\gamma \in (\gamma_2, \hat{\gamma})$ such that $\Pi(\delta, \frac{1-F(\gamma)}{1-F(\gamma)+F(\gamma)\rho(\gamma)}, \gamma) = 0$. As a result, the minimum equilibrium cutoff in the communication game is strictly lower than that in the simultaneous game for any strictly positive honesty rate. This also implies that the rate of coordination on cooperation and the continuation game cooperation rate in the communication game are strictly higher than the corresponding rates in the simultaneous game.

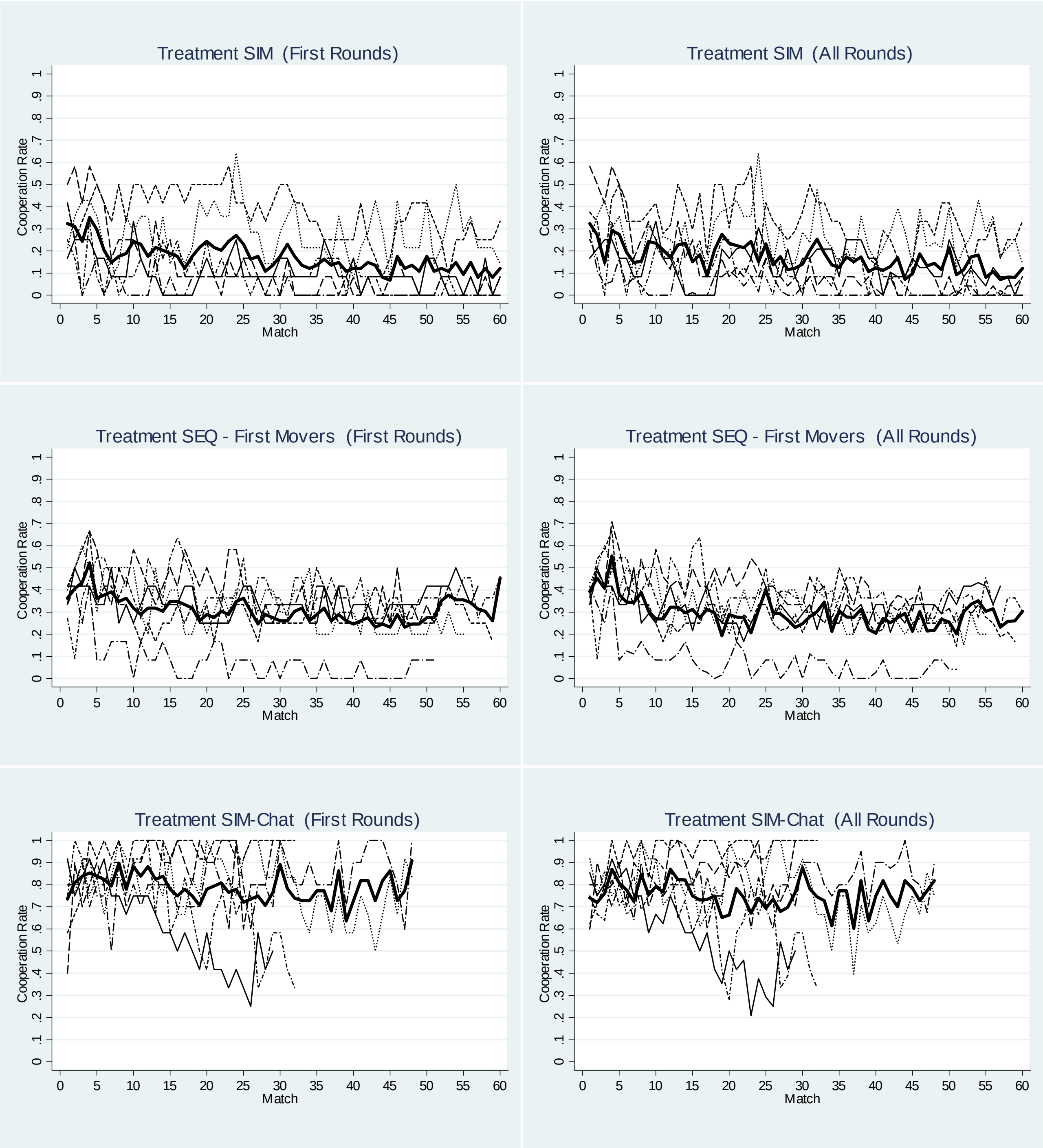
We will now derive an upper bound on the dishonesty rate so that the statement in Proposition 3 regarding the comparison between the communication game and the sequential game holds. Let γ_1^* and γ_2^* denote the sequential game equilibrium cutoff for the first mover and the second mover, respectively as in the proof of Proposition 2 above. Let $\hat{\gamma}$ solve $(1 - F(\hat{\gamma}))^2 = (1 - F(\gamma_1^*))(1 - F(\gamma_2^*))$. As $\gamma_2^* < \gamma_1^*$, $\hat{\gamma} \in (\gamma_2^*, \gamma_1^*)$ must hold. Next note that there exists a number $\hat{\rho} \in (0, 1)$ such that $\Pi(\delta, \frac{1-F(\hat{\gamma})}{1-F(\hat{\gamma})+F(\hat{\gamma})\hat{\rho}}, \hat{\gamma}) = 0$ because if $\hat{\rho} = 0$, then $\Pi(\delta, 1, \hat{\gamma}) > 0$ (as $\hat{\gamma} > \gamma_2^*$) and if $\hat{\rho} = 1$, then $\Pi(\delta, 1 - F(\hat{\gamma}), \hat{\gamma}) < 0$ (as $\hat{\gamma} < \gamma_1^*$, and $\Pi(\delta, 1 - F(\gamma), \gamma) < 0$ for every $\gamma < \gamma_1^*$ due to A2, Claim 1 and the fact that γ_1^* is strictly lower than the minimum equilibrium cutoff in the simultaneous game γ^* which solves $\Pi(\delta, 1 - F(\gamma^*), \gamma^*) = 0$). Thus, for every honesty function

$\rho(\gamma)$ such that $\rho(\gamma) < \hat{\rho}$, the statement in Proposition 3 regarding the comparison between the communication game and the sequential game holds because there exists a $\tilde{\gamma}$ such that $\tilde{\gamma} < \hat{\gamma}$ and $\Pi(\delta, 1 - F(\tilde{\gamma})\rho(\tilde{\gamma}), \tilde{\gamma}) = 0$. Such $\tilde{\gamma}$ exists because (i) $\Pi(\delta, \frac{1-F(\hat{\gamma})}{1-F(\hat{\gamma})+F(\hat{\gamma})\rho(\hat{\gamma})}, \hat{\gamma}) > 0$ given Claim 1 and $\rho(\hat{\gamma}) < \hat{\rho}$, and (ii) there must exist $\gamma < \hat{\gamma}$ such that $\Pi(\delta, \frac{1-F(\gamma)}{1-F(\gamma)+F(\gamma)\rho(\gamma)}, \gamma) < 0$ by A2 and Claim 1. As a result, there exists $\tilde{\gamma} < \hat{\gamma}$ such that $\Pi(\delta, \frac{1-F(\tilde{\gamma})}{1-F(\tilde{\gamma})+F(\tilde{\gamma})\rho(\tilde{\gamma})}, \tilde{\gamma}) = 0$, and $\tilde{\gamma}$ is an equilibrium cutoff of the communication game. It follows that $(1 - F(\tilde{\gamma}))^2 > (1 - F(\gamma_1^*))(1 - F(\gamma_2^*))$.

Online Appendix (Not for Publication)

A Additional Figures and Tables

Figure 3: Evolution of Cooperation in Treatments and Sessions



Notes: Only first-mover data for treatment SEQ. The left (right) panels uses data from first (all) rounds of all matches. Thin (bold) lines indicate matching group (treatment) averages.

Table 7: Summary statistics on numbers of messages sent

Round	Mean	Median	Minimum	Maximum	Frequency
1	2.23	2	0	7	2,456
2	1.59	2	0	6	1,310
3	1.47	1	0	7	704
4	1.43	1	0	8	430
5	1.24	1	0	5	162
6	1.63	2	0	4	70
7	1.32	1	0	3	22
8	1.41	1	0	3	22

Note: Data from all matches and rounds.

B Treatment and session information

Table 8: Treament and session information

Treatment	Matching Group	# Subjects	# of games played	Avg. percentage of mutual cooperation rate
SIM	1	12	60	3.26
SIM	2	12	60	3.26
SIM	3	14	60	8.43
SIM	4	12	60	0.14
SIM	5	12	60	11.53
SIM	6	12	60	1.36
SEQ	1	24	57	24.31
SEQ	2	24	52	26.94
SEQ	3	20	55	18.48
SEQ	4	24	51	4.17
SEQ	5	24	59	17.74
SEQ	6	22	60	27.51
CHAT	1	12	29	46.77
CHAT	2	10	29	74.52
CHAT	3	12	48	64.06
CHAT	4	10	48	72.71
CHAT	5	12	32	89.91
CHAT	6	12	32	50.00

Note: In treatment SEQ, subjects are assigned to one of two player roles (first mover or second mover). Therefore, matching groups for Treatment SEQ had to be larger for reasons of comparability.

C Projection of mutual cooperation levels over a long time horizon

In this Appendix we assess to what levels cooperation and mutual cooperation rates might have converged to had the experiment been conducted over a very long time horizon. For this purpose, we employ an approach similar to the one suggested in Noussair et al. (1995) and Barut *et al.* (2002). For instance, to estimate and statistically compare asymptotes of cooperation rates in treatments SIM (all data) and SEQ (first-mover data) we run the following two OLS regressions:

Table 9: Estimated asymptotes and test results

Cooperation Rates

SEQ		**SIM**		**CHAT**
0.269	$>^{***}$	0.169	$<^{***}$	0.744
(0.005)		(0.013)		(0.031)

Mutual Cooperation Rates

SEQ		**SIM**		**CHAT**
0.176	$>^{***}$	0.045	$<^{***}$	0.666
(0.006)		(0.006)		(0.039)

Notes: The table shows the estimated asymptotes from equations of the form (2) and (3). Standard errors in parentheses. The "<" and "≈" signs summarize the results of statistical tests based on the z-statistic defined in equation (4), with "≈" indicating no significant difference at the 10%-level. The superscripts indicate the level of significance, where *** and ** indicates significance at the 1% and 5% level, respectively.

$$\text{Coop}_{ijt} = \sum_{j=1}^{6} \alpha_j^{SIM} \times \frac{D_j^{SIM}}{t} + \beta^{SIM} \times \frac{(t-1)}{t} + \varepsilon \tag{2}$$

and

$$\text{Coop}_{ijt} = \sum_{j=1}^{6} \alpha_j^{SEQ} \times \frac{D_j^{SEQ}}{t} + \beta^{SEQ} \times \frac{(t-1)}{t} + \varepsilon, \tag{3}$$

where Coop_{ijt} denotes the observed average cooperation rate in group i, of matching group j in match t; D_j^{SIM} (D_j^{SEQ}) is a dummy for matching group j of treatment SIM (SEQ). The regressions were run with all data and with observations clustered at the matching group level.

Following the interpretation in Barut *et al.* (2002), the coefficient α_j^{SIM} (α_j^{SEQ}) is an estimate of the average cooperation rate in match 1 of matching group j in treatment SIM (SEQ), whereas β^{SIM} (β^{SEQ}) is the estimated average cooperation rate to which the time series converge if $t \rightarrow \infty$. Hence, this model allows the starting cooperation rates to be different across the individual matching groups of a treatment but assumes the cooperation rates in all matching groups of a treatment to converge to a common asymptote. (Similar regressions were run to estimates asymptotes for mutual cooperation rates.)

We are interested in whether in the very long run the cooperation rates across treatments would have converged to the same level. To test the hypothesis H_0: $\beta^{SIM} = \beta^{SEQ}$, we compute the statistic

$$z = \frac{\widehat{\beta}^{SIM} - \widehat{\beta}^{SEQ}}{\sqrt{se(\widehat{\beta}^{SIM})^2 + se(\widehat{\beta}^{SEQ})^2}}, \tag{4}$$

where $\widehat{\beta}^{SIM}$ and $\widehat{\beta}^{SEQ}$ are the estimated coefficients and $se(\widehat{\beta}^{SIM})$ and $se(\widehat{\beta}^{SEQ})$ the standard errors from equations (2) and (3), respectively. Table 9 shows the results. The "<" and ">" signs in Table 9 summarize the results of statistical tests based on the z-statistic defined in (4). The superscripts indicate the level of significance. We find that all asymptotes are precisely estimated to be significantly larger than 0. Furthermore, the estimated asymptotes for treatments SEQ and CHAT, respectively, are significantly larger than the estimated asymptotes for treatment SIM. To the extent that the above projection technique is justified, these results mean that the treatment effects reported in Tables 1 and 2 would have prevailed in case the experiments had consisted of many more matches.

D Summary of chat contents

Table 10: Summary of chat contents

Verbal Code	Explanation	N	%	Cum %
both1	Subject suggests that both play 1	1,278	32.80	32.80
agree	Subject agrees to a proposal	617	15.84	48.64
none	Subject does not communicate	506	12.99	61.63
trivial	Small talk, off topic	422	10.83	72.46
same	Subject suggests/announces to play as before	282	7.24	79.70
both1 always	Subject suggests that both always play 1	158	4.06	83.75
self1	Subject announces to play 1	158	4.06	87.81
both1 again	Subject suggests that both play 1 again	114	2.93	90.73
complain	Subject complains about behavior of other	86	2.21	92.94
both2	Subject suggests that both play 1	57	1.46	94.40
greeting	Subject greets the other	47	1.21	95.61
self2	Subject announces to play 2	29	0.74	96.36
what to do?	Subject asks what to do	28	0.72	97.07
sorry	Subject apologizes for own behavior	21	0.54	97.61
unclear	Subject answers but it is not clear if she (dis)agrees	21	0.54	98.15
me1 you2	Subject suggests that she plays 1, the other 2	18	0.46	98.61
other1	Subject asks the other player to choose 1	18	0.46	99.08
disagree	Subject disagrees with a proposal or statement	10	0.26	99.33
me2 you1	Subject suggests that she plays 2, the other 1	9	0.23	99.56
weak agree	Subject weakly agrees to a proposal	8	0.21	99.77
alternating me1 you2	Subject suggests to alternate in choices	4	0.10	99.87
same always	Subject suggests that both play always the same action	3	0.08	99.95
question	Subject asks a question	1	0.03	99.97
risk	Subject mentions riskiness w.r.t. own or other action	1	0.03	100.00
Total		3,896	100.00	

Notes: Data from all rounds of all matches for which the verbal codes assigned to chats by both coders coincided. Note that in the instructions the cooperative (defective) action was labeled 1 (2).

E Experimental Instructions

On the next pages, we reproduce the experimental instructions used in treatment SIM, SEQ, and CHAT, respectively.

Instructions (SIM)

Welcome to this experiment!

You are about to participate in an experiment on decision-making, and you will be paid for your participation in cash, privately at the end of the experiment. What you earn depends partly on your decisions, partly on the decisions of others, and partly on chance.

Please turn off your cellular phone now.

The entire experiment will take place through computer terminals, and all interaction between you will take place through the computers. Please do not talk or in any way try to communicate with other participants during the experiment. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you.

General Instructions

1. In this experiment you will be asked to make decisions in several rounds. You will be randomly paired with another participant for a sequence of rounds. Each sequence of rounds is referred to as a match.

2. The length of a match is randomly determined. After each round of a match, there is a 50% chance that the match will continue for at least another round. For this purpose, at the end of each round of a match the computer will roll a fair 100-sided die and the match will continue if the die shows a number between 1 and 50 and the match will end if the die shows a number between 51 and 100. So, for instance, if you are in round 1 of a match, the chance that there will be a 2nd round is 50%, and if you are in round 2 of a match, the chance that there will be a 3rd round is also 50%, and so on.

3. Once a match ends, you will be randomly paired with another participant for a new match.

4. The choices and the payoffs associated with each choice in each round are as follows:

	the other’s choice	
your choice	1	2
1	32, 32	12, 50
2	50, 12	25, 25

The first entry in each cell represents your payoff, while the second entry represents the payoff of the participant you are matched with.

Once you and the participant you are paired with have made your choices, those choices will be highlighted and your payoff for the round will appear.

Payoffs are as indicated in the table above. That is, if:
You select 1 and the other selects 1, you each make 32.
You select 1 and the other selects 2, you make 12 while the other makes 50.
You select 2 and the other selects 1, you make 50 while the other makes 12.
You select 2 and the other selects 2, you each make 25.

The experiment will end after the first match that is completed after 75 minutes have passed, or after 60 matches have been completed, whichever is the sooner.

At the end of the experiment you will be paid € 0.007 for every point scored.

Before the start of the experiment, let us remind you that:

- The length of a match is randomly determined. After each round, there is a 50% chance that the match will continue for at least another round. You will interact with the same participant for the entire match.
- After a match is finished, you will be randomly paired with another participant for a new match.

Instructions (SEQ)

Welcome to this experiment!

You are about to participate in an experiment on decision-making, and you will be paid for your participation in cash, privately at the end of the experiment. What you earn depends partly on your decisions, partly on the decisions of others, and partly on chance.

Please turn off your cellular phone now.

The entire experiment will take place through computer terminals, and all interaction between you will take place through the computers. Please do not talk or in any way try to communicate with other participants during the experiment. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you.

General Instructions

1. In this experiment you will be asked to make decisions in several rounds. You will be randomly paired with another participant for a sequence of rounds. Each sequence of rounds is referred to as a match.

2. The length of a match is randomly determined. After each round of a match, there is a 50% chance that the match will continue for at least another round. For this purpose, at the end of each round of a match the computer will roll a fair 100-sided die and the match will continue if the die shows a number between 1 and 50 and the match will end if the die shows a number between 51 and 100. So, for instance, if you are in round 1 of a match, the chance that there will be a 2nd round is 50%, and if you are in round 2 of a match, the chance that there will be a 3rd round is also 50%, and so on.

3. Once a match ends, you will be randomly paired with another participant for a new match.

4. The choices and the payoffs associated with each choice in each round are as follows:

	the other's choice	
your choice	1	2
1	32, 32	12, 50
2	50, 12	25, 25

The first entry in each cell represents your payoff, while the second entry represents the payoff of the participant you are matched with.

Once you and the participant you are paired with have made your choices, those choices will be highlighted and your payoff for the round will appear.

At the beginning of the experiment each participant will be randomly assigned the role of participant A or participant B. You will be informed about your role when the experiment starts. You will keep this role for the entire experiment.

In each round of a match, participant A makes his/her decision first. Then, after observing the decision of participant A, participant B makes his/her decision. (The decision of A will be highlighted on the screen of B when B makes his/her decision.)

Payoffs are as indicated in the table above. First, assume that you are participant A. Then, you make the first decision. If:
You select 1 and participant B selects 1, you each make 32.
You select 1 and participant B selects 2, you make 12 while participant B makes 50.
You select 2 and participant B selects 1, you make 50 while participant B makes 12.
You select 2 and participant B selects 2, you each make 25.

Next, assume that you are participant B. Then, you make a decision after observing the decision of participant A. If:
Participant A selects 1 and you select 1, you each make 32.
Participant A selects 1 and you select 2, you make 50 while participant A makes 12.
Participant A selects 2 and you select 1, you make 12 while participant A makes 50.

Participant A selects 2 and you select 2, you each make 25.

The experiment will end after the first match that is completed after 75 minutes have passed, or after 60 matches have been completed, whichever is the sooner.

At the end of the experiment you will be paid € 0.007 for every point scored.

Before the start of the experiment, let us remind you that:

- The length of a match is randomly determined. After each round, there is a 50% chance that the match will continue for at least another round. You will interact with the same participant for the entire match.
- After a match is finished, you will be randomly paired with another participant for a new match.
- In each round of a match, participant A makes his/her decision first. Then, after observing the decision of participant A, participant B makes his/her decision.

Instructions (CHAT)

Welcome to this experiment!

You are about to participate in an experiment on decision-making, and you will be paid for your participation in cash, privately at the end of the experiment. What you earn depends partly on your decisions, partly on the decisions of others, and partly on chance.

Please turn off your cellular phone now.

The entire experiment will take place through computer terminals, and all interaction between you will take place through the computers. Please do not talk or in any way try to communicate with other participants during the experiment. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you.

General Instructions

1. In this experiment you will be asked to make decisions in several rounds. You will be randomly paired with another participant for a sequence of rounds. Each sequence of rounds is referred to as a match.

2. The length of a match is randomly determined. After each round of a match, there is a 50% chance that the match will continue for at least another round. For this purpose, at the end of each round of a match the computer will roll a fair 100-sided die and the match will continue if the die shows a number between 1 and 50 and the match will end if the die shows a number between 51 and 100. So, for instance, if you are in round 1 of a match, the chance that there will be a 2nd round is 50%, and if you are in round 2 of a match, the chance that there will be a 3rd round is also 50%, and so on.

3. Once a match ends, you will be randomly paired with another participant for a new match.

4. The choices and the payoffs associated with each choice in each round are as follows:

	the other’s choice	
your choice	1	2
1	32, 32	12, 50
2	50, 12	25, 25

The first entry in each cell represents your payoff, while the second entry represents the payoff of the participant you are matched with.

Once you and the participant you are paired with have made your choices, those choices will be highlighted and your payoff for the round will appear.

Payoffs are as indicated in the table above. That is, if:
You select 1 and the other selects 1, you each make 32.
You select 1 and the other selects 2, you make 12 while the other makes 50.
You select 2 and the other selects 1, you make 50 while the other makes 12.
You select 2 and the other selects 2, you each make 25.

During a match a dialogue box is available in which you can exchange messages with the participant you are paired with. Although we will record these messages, only you and the participant you are paired with in a match will see them. Think of the dialogue box as your own private dialogue system to help you decide what to do. Note, in sending messages back and forth between you and the participant you are paired with in a match we request you follow three simple rules: (1) Discussion must be in English. No other language is allowed. (2) Be civil to each other, don’t use bad language, and don’t make any threats to each other. (3) Do not identify yourself, your seat number or anything that might reveal your identity. The dialogue box is intended for you to use to discuss your choices and should be used that way. During the first 5 matches you are able to exchange messages for 30 seconds per round; as of match 6 for 15 seconds per round.

The experiment will end after the first match that is completed after 75 minutes have passed, or after 60 matches have been completed, whichever is the sooner.

At the end of the experiment you will be paid € 0.007 for every point scored.

Before the start of the experiment, let us remind you that:

- The length of a match is randomly determined. After each round, there is a 50% chance that the match will continue for at least another round. You will interact with the same participant for the entire match.
- After a match is finished, you will be randomly paired with another participant for a new match.