Center for Mathematical Economics
Working Papers

652

August 2021

Inequity Aversion and Limited Foresight in the Repeated Prisoner's Dilemma

Teresa Backhaus and Yves Breitmoser

Center for Mathematical Economics (IMW)
Bielefeld University
Universitätsstraße 25
D-33615 Bielefeld · Germany

e-mail: imw@uni-bielefeld.de
uni-bielefeld.de/zwe/imw/research/working-papers
ISSN: 0931-6558

Inequity Aversion and Limited Foresight in the Repeated Prisoner's Dilemma

Teresa Backhaus*
FU Berlin and WZB

Yves Breitmoser
Bielefeld University

August 20, 2021

Abstract

Reanalyzing 12 experiments on the repeated prisoner's dilemma (PD), we robustly observe three distinct subject types: defectors, cautious cooperators and strong cooperators. The strategies used by these types are surprisingly stable across experiments and uncorrelated with treatment parameters, but their population shares are highly correlated with treatment parameters. As the discount factor increases, the shares of defectors decrease and the relative shares of strong cooperators increase. Structurally analyzing behavior, we next find that subjects have limited foresight and assign values to all states of the supergame, which relate to the original stage-game payoffs in a manner compatible with inequity aversion. This induces the structure of coordination games and approximately explains the strategies played using Schelling's focal points: after (c,c) subjects play according to the coordination game's cooperative equilibrium, after (d,d) they play according to its defective equilibrium, and after (c,d) or (d,c) they play according to its mixed equilibrium.

JEL–Code: C72, C73, C92, D12
Keywords: Repeated game, Behavior, Tit-for-tat, Mixed strategy, Memory, Belief-free equilibrium, Laboratory experiment

*We thank Andrea Ariu, Kai Barron, Niels Boissonnet, Friedel Bolle, Fabian Dvorak, Sebastian Fehrler, Drew Fudenberg, Daniel Gietl, Paul Heidhues, Steffen Huck, Gustav Karreskog, Johannes Leutgeb, Vlada Pleshcheva, Sebastian Schweighofer-Kodritsch, Roland Strausz, Robert Stüber, Georg Weizsäcker, Roel van Veldhuizen, and Joachim Winter, as well as the audiences in Schwanenwerder, Florence, Tel Aviv, Konstanz, and Berlin for many helpful comments. Financial support by the Deutsche Forschungsgemeinschaft (BR 4648/1 and CRC TRR 190) is gratefully acknowledged. Teresa Backhaus: Wissenschaftszentrum Berlin für Sozialforschung, Reichpietschufer 50, 10785 Berlin, Germany, email: teresa.backhaus@wzb.eu, phone: +49 30 25491 411. Yves Breitmoser (corresponding author): Universitätsstr. 25, 33615 Bielefeld, Germany, email: yves.breitmoser@uni-bielefeld.de.

1 Introduction

One of the most dynamic research fields over the last two decades has been behavioral game theory, i.e. the econometric and theoretical analysis of laboratory games to align observed behavior with game-theoretical concepts. How should we think of beliefs, utilities, and subjects' choices, and is it possible to explain choices as responses to incentives? In some classes of games, most notably auctions, behavior shows to be reasonably consistent with theory after simply accounting for risk aversion (Bajari and Hortacsu, 2005) or biased beliefs (Eyster and Rabin, 2005). In generic normal-form games involving dominated strategies, behavior is captured after relaxing rational expectations (Costa-Gomes et al., 2001); in games without dominated strategies, behavior tends to mainly reflect logistic errors in choice (Weizsäcker, 2003; Brunner et al., 2011); and in games involving the distribution of monetary benefits, preference interdependence seems to organize behavior (Fehr and Schmidt, 1999; Charness and Rabin, 2002). Particular behavioral models tend to be disputed, but overall, there has been substantial progress in aligning observed behavior and theoretical predictions across many classes of games.

One class of games that has experienced less progress in aligning behavior and predictions is the large class of repeated games. Repeated games are the main approach toward modeling long-run interactions, in particular to study cooperation and defection, and they have been a core object of game-theoretic analyses at least since the folk theorem for repeated games with discounting (Fudenberg and Maskin, 1986). Regarding behavior in experiments, however, there is no consensus on what subjects actually do—not even whether they play pure, mixed or behavior strategies—and there is not a single analysis relating round-by-round decisions to beliefs and expected utilities despite its common practice in structural analyses of behavior in static games.

The purpose of this paper is to re-analyze a large data set comprising 12 experiments to robustly estimate strategies and structurally analyze them similar to previous work on static games. We seek to answer three questions: Which strategies do subjects actually play? Are the strategies played and the shares of them predictable across conditions? How do the strategies align with expected utilities, and to what extent is individual behavior consistent with existing models of behavior in games? Regarding the first question, much of the existing literature restricts attention to strategies that are pure (with some noise), but recent evidence suggests that behavior strategies might better explain behavior (reviewed below). Regarding the second question, existing evidence suggests that the type shares playing specific cooperative strategies fluctuate fairly erratically between treatments, which is puzzling but may reflect inadequate constraints to pure strategies that we shall relax in our analysis. The third question is novel in the analysis of repeated games and has been left unexplored in previous work, but it is a central question in many behavioral papers and of major relevance in order to link behavior in repeated games to other literature.

Our main results can be summarized as follows. Estimating the individual strategies without restrictions to pure or otherwise known strategies, and across 145,000 decisions from 12 experiments, we find that both the bottom-up and the top-down approach toward behavioral modeling imply that there are three subject types robustly observed across all

treatments and experiments. Type 1 plays a slightly perturbed version of always defect. We refer to subjects of type 1 as defectors. Types 2 and 3 both play behavior strategies predicting nearly pure behavior after (c,c) and (d,d) (cooperation and defection, respectively), and they both randomize after (c,d) and (d,c). In round 1, however, type 2 cooperates with intermediate probability and type 3 with high probability. We refer to subjects of types 2 and 3 as cautious and strong cooperators, respectively. The states will be abbreviated as cc, cd, dc, dd in the following. We obtain these results from an unrestricted estimation of memory-1 strategies, i.e. an estimation of strategies without imposing restrictions to say pure strategies that characterize previous work, and we examine robustness to memory-2 in the appendix. This unrestricted estimation is a first major novelty of our analysis and addresses the first question above.

Further, as a second major novelty, we use data-mining techniques to obtain an upper bound for the goodness-of-fit that could be obtained assuming all subjects play versions of pure strategies. We relax many assumptions made in the literature, grant many degrees of freedom "for free", and allow for either no switching, random switching, or Markov switching of strategies between supergames. This generality notwithstanding, the upper bound for pure strategies is significantly lower than the goodness-of-fit of the simple three-type model described above identified by unrestricted estimation—which rules out that cooperating subjects (i.e. the ones not playing always defect) are well-described as playing pure strategies.

Across treatments, the three types of strategies used are largely uncorrelated with treatment parameters or other known predictors of cooperation, while the distribution of types is highly correlated with the discount factor δ: As δ approaches the Blonski et al. (2011) (BOS) threshold of cooperation δ^*, the share of defectors decreases relative to cooperators, and as δ is raised further, the strong cooperators start to outnumber the cautious cooperators. That is, as a third major novelty, the unrestricted estimation implies that the distribution of the strategies becomes predictable—addressing the second question above—but the types of strategies they play are all the more puzzling. Specifically, we have no prior explanation for the observation that type shares are correlated with δ (in relation to the BOS threshold δ^*, see Figure 4), which suggests that subjects are aware of δ and other parameters when choosing their strategy, while the actual strategies are largely uncorrelated with δ (Figure 3).

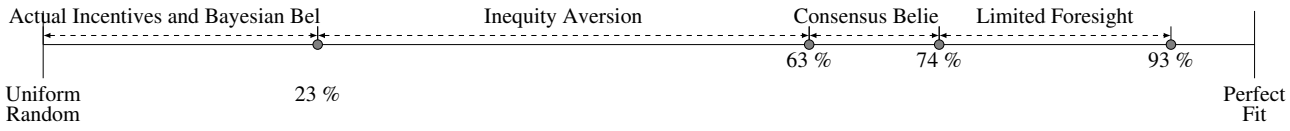
To resolve this puzzle, and the third question above, we use techniques developed for the estimation of static games (McKelvey and Palfrey, 1995; Bajari and Hortacsu, 2005) and dynamic games (Aguirregabiria and Mira, 2007) in order to understand the individual motivation behind the strategies of cooperative types across treatments. In this way, we also seek to resolve a second puzzle that the above results highlight: Cooperating subjects in our unrestricted analysis, and indeed in all previous work, cooperate with a probability close to 1 if both subjects had cooperated in the previous round. They do so even if the expected payoff of cooperating (in the next round) is substantially below the expected payoff of defecting, as we demonstrate below, and even if Grim is not a subgame perfect equilibrium. The latter implies that behavior cannot be explained just by relaxing beliefs about the opponent's strategy. By estimating the dynamic games, we try to understand subjects' preferences in a manner similar to previous behavioral analyses of cooperative subjects in one-shot games, which is the fourth major novelty of our analysis.

We find that the strategies of cooperators are consistent with false consensus beliefs about the opponent's type, i.e. that each subject believes their opponent to be of the same type as they are (as in symmetric equilibrium), that subjects display severely limited foresight (as if the discount factor was zero, discussed shortly), and that their preferences are well described by Fehr-Schmidt inequity aversion. The limited foresight implies that subjects do not look beyond the outcome of the present round and do not explicitly consider sums of discounted payoffs. Instead, subjects associate utility values with each of the four possible outcomes of the present round (cc, dc, cd, dd) that encode the subject's value of reaching the respective state. We estimate that these state values induce a coordination game played round-by-round, that is with one round of foresight, and that these state values can be derived from the stage game payoffs using inequity aversion.

As usual, this coordination game has three Nash equilibria: a cooperative one, a defective one, and a mixed one. Our analysis indicates that, reminiscent of Schelling's focal points, after *cc* subjects expect cooperation (i.e. believe the opponent to cooperate) and play the cooperative equilibrium of the coordination game, after *dd* they expect defection and play the defective equilibrium, and after mixed histories (*cd* or *dc*) they play the mixed equilibrium. Our results indicate a similar line of reasoning in round 1: Given the actual treatment parameters, some subjects focus on the cooperative equilibrium (the "strong cooperators"), some focus on the defective equilibrium (the "defectors"), and some seem "unsure" playing the mixed equilibrium in round 1 (the "cautious cooperators"). The respective subject shares are predictable using the distance of discount factor δ to the BOS-threshold δ^* derived by Blonski et al. (2011).

We thus obtain a closed behavioral foundation of the strategies played in the repeated PD. In contrast to previous work, this model *explains* the strategies that we estimated *without restrictions* beyond standard memory-1, rather than merely estimating strategy weights under non-validated restrictions to certain pure strategies. Further, our results bear many relations to previous work in behavioral economics.

Figure 1: Decomposition of behavior into model components
(second halves of sessions, based on 79.892 observations)



Note: This plot summarizes the results of our structural analyses (Table 5) of the strategies played. Here we focus on the strategies of subjects not classified as playing always defect. The clairvoyance model explaining the strategies of cooperating subjects perfectly across treatments obtains the "perfect fit" (100%), while the model predicting uniform randomization obtains 0%. The remaining percentages are computed proportionally to these two benchmarks. First, the model assuming that subjects play logit responses to Bayesian beliefs over their opponent's strategy obtains 23%, additionally allowing for inequity aversion increases the score to 63%, next adopting consensus beliefs increases the explained variance to 74%, and allowing for a flexible discount factor δ (leading to limited foresight as an estimation result) captures an additional 19% for a total of 93%. Figure 6 provides information also for behavior in the first halves of sessions and on robustness with respect to modeling assumptions.

Quantitatively, using just four parameters for the 80.000 observations of "experienced subjects" (in their second halves of sessions), the resulting model captures 93% of observed variance across 32 treatments from 12 experiments. Figure 1 briefly summarizes the obtained decomposition of behavior using our structural estimates. Our finding that behavior in the repeated PD is characterized by false consensus beliefs relates to a central concept in psychology (Ross et al., 1977) implying symmetric equilibrium; limited foresight relates to a central concept in computer science and behavioral game theory (Jehiel, 2001; Kübler and Weizsäcker, 2004); and inequity aversion is a central concept in behavioral economics (Fehr and Schmidt, 1999).[1] Further, the idea that a repeated PD resembles a coordination game in round 1, and iteratively in any subsequent round, has been discussed at least since Rabin (1993). We provide the first empirical confirmation, by demonstrating that this implicit coordination game is endogeneously obtained as an estimation result using econometric techniques known from static games, and by the findings that this coordination game is highly predictable using inequity aversion and that its equilibria are highly predictive of behavior across treatments and experiments. This yields a first explanation for behavior in repeated games and a first set of precise behavioral predictions for future work on repeated games generalizing the repeated PD—a very encouraging step to bridge the gap between behavioral analyses of repeated games and behavioral analyses of static games.

2 Background information

Definitions The *prisoner's dilemma* (PD) involves two players choosing whether to cooperate (c) or defect (d). In the normalized PD, each player's payoff is 1 if both cooperate and 0 if both defect. If exactly one player cooperates, the cooperating player's payoff is $-l$ ($l > 0$) and the defecting player's payoff is $1 + g$ ($g > 0$). An infinite repetition of this constituent game is strategically equivalent to an indefinitely repeated one that is terminated with probability $1 - \delta$ after each round, assuming players are risk neutral and discount future payoffs exponentially (using factor $\delta < 1$). We will refer to these games jointly as repeated PD (or, *supergame*). Given $g, l > 0$, cooperation is dominated in the one-shot game but may be sustained along the path of play in subgame-perfect equilibria of the repeated PD (depending on δ).

A *strategy* σ in the repeated PD maps all finite histories to probabilities of cooperation in the next round. The strategy has *memory-1* if it prescribes the same cooperation probability for any two histories not differing in the actions chosen in their respective last rounds. It has memory-2 if the same holds for the last two rounds. We denote memory-1 strategies as $\sigma = (\sigma_\emptyset, \sigma_{cc}, \sigma_{cd}, \sigma_{dc}, \sigma_{dd})$ corresponding to the five memory-1 histories $\{\emptyset, cc, cd, dc, dd\}$, called *states* in the following. For example, σ_{cd}, denotes the probability of cooperation when a player's most recent action is c and her opponent's most recent action is d. A strat-

[1] Interestingly, our results shed new light on the findings of Dreber et al. (2014), who found that inequity aversion does not help explain behavior in the repeated PD. The difference in our analysis is that we do not attempt to explain so-called standard strategies, i.e. pure strategies, but behavior strategies estimated without restrictions to purity. Further, inequity aversion in our analysis is closely interlinked with limited foresight, which Dreber et al. (2014) did not include in their analysis, and we allow for α and β to be free parameters.

egy is a *pure strategy* if it prescribes degenerate cooperation probabilities after all histories ($\sigma \in \{0,1\}^5$), and it is a *behavior strategy* otherwise. It is a *mixed strategy*, when a player randomizes over the set of pure strategies prior to the start of each supergame, but sticks to the drawn pure strategy throughout the supergame. In contrast, when playing a behavior strategy, she randomizes during the supergame.[2]

Table 1: Overview of most commonly analyzed strategies (see Table 12 in the appendix for a more comprehensive list)

Strategy	Abbreviation	Description	$(\sigma_\emptyset, \sigma_{cc}, \sigma_{cd}, \sigma_{dc}, \sigma_{dd})$
Always Defect	AD	Always defects	(0,0,0,0,0)
Always Cooperate	AC	Always cooperates	(1,1,1,0,0)
Grim	G	Only cooperate in R1 and after cc	(1,1,0,0,0)
Tit-for-Tat	TFT	Start with c, then copy opponent	(1,1,0,1,0)
Suspicious TFT	STFT, D-TFT	Start with d, then copy opponent	(0,1,0,1,0)
Win-Stay-Lose-Shift	WSLS	Cooperate in R1, cc and dd	(1,1,0,0,1)
Semi-Grim	SG	Behavior strategy satisfying ...	$\sigma_{cd} = \sigma_{dc}$

Note: The conventional definition of AC is $(1,1,1,1,1)$, which is behaviorally equivalent to $(1,1,1,0,0)$. The definition used above implies that any memory-1 behavior strategy that might be observed on average can be rebuilt using some combination of AD, AC, Grim, TFT and WSLS.

Related behavioral literature We will keep the literature review short and focused due to the availability of an excellent recent survey by Dal Bó and Fréchette (2018). The modern experimental research on the repeated PD started with Dal Bó (2005), who criticized earlier experiments for implementing experimental designs that let subjects play against computerized opponents. The first wave of experiments following Dal Bó (2005) includes Dreber et al. (2008), Duffy and Ochs (2009), Blonski et al. (2011) and Kagel and Schley (2013), and focuses on analyzing first-round and total cooperation rates. A second wave comprising Dal Bó and Fréchette (2011, 2015), Bruttel and Kamecke (2012), Camera et al. (2012), Fudenberg et al. (2012), Sherstyuk et al. (2013), Breitmoser (2015), and Fréchette and Yuksel (2017) analyzes the strategies actually chosen by players. The general theme in the reported results is that initial cooperation rates depend on the strategic environment. More specifically, the results indicate that subgame perfection of Grim is necessary but not sufficient for cooperation to emerge (first reported in Dal Bó, 2005), and that subsequent cooperation of subjects depends on their opponent's actions, primarily on those in the previous round. The central importance of initial cooperation is also demonstrated in Fudenberg and Karreskog (2020). Many of the second-wave analyses classify individual subjects' strategies into varying sets of pre-selected strategies. Even allowing for noise, these analyses clearly show that subjects do not homogeneously follow a given pure strategy across all supergames. The studies differ in their assumptions of what subjects might be doing instead—whether they are playing pure, mixed, or behavior strategies—and consequently in their conclusions about behavior.

Most analyses assume that decisions are made only prior to the first supergame of a session, with subjects then sticking to a *pure strategy* for the rest of the session. Given this

[2]Including the case when she would switch between pure strategies within a supergame.

restriction to pure strategies, these analyses typically conclude that the majority of subjects play either AD, TFT, or Grim, with each being attributed weights around 20–30%. For example, Result 6 of Dal Bó and Fréchette (2018, DF18) states that these three strategies account for "most of the data", specifically they "account for 70 percent of strategies in most treatments", but importantly, this result is obtained after a-priori restricting attention to (a subset of) pure strategies without further validating this restriction. We refer to this statement as the *pure-strategy conjecture*.

A second, less common approach is based on the assumption that subjects switch pure strategies between supergames, which we refer to as *mixed strategies* in the game-theoretical sense. For example, DF18 report that 84 percent of choices in supergames lasting more than one round are accounted for by five pure strategies (now also including AC and suspicious TFT) when they allow for strategy switching between supergames (DF18, Footnote 38).[3] The difficulty now is to explain this strategy switching; otherwise, the impression of a perfect fit, not requiring a complicated analysis allowing for noise, is intriguing, but it is only true in a post-hoc sense. Ex-ante, the strategy chosen by a given subject is not perfectly predictable, and the game-theoretical concept closest to such a random choice over varying pure strategies over time is that of a mixed strategy. The probabilities of choosing different pure strategies over time may be path dependent, given the path-dependency they may be degenerate, and they may be heterogeneous between subjects. Below, we shall explicitly allow for these possibilities by considering Markov-switching models to capture strategy switching that contain pure, mixed, and path-dependent mixtures as special cases. This will be one of the major novelties of our analysis and will enable us to determine an upper bound for the goodness-of-fit of pure and mixed strategies.

A third and growing group of studies challenges the pure-strategy conjecture by allowing subjects to randomize in each round of each supergame, as in the game-theoretical concept of *behavior strategies*. Relaxing the restriction to pure strategies, Breitmoser (2015) observed that cooperating subjects play a semi-grim behavior strategy $(\sigma_\emptyset, \sigma_{cc}, \sigma_{cd}, \sigma_{dc}, \sigma_{dd})$ satisfying $\sigma_{cc} > \sigma_{cd} \approx \sigma_{dc} > \sigma_{dd}$ and $\sigma_\emptyset \in [0,1]$ (*behavior-strategy conjecture*).[4] Additionally, semi-grim behavior strategies are found to better capture behavior than mixtures of pure memory-1 strategies.[5] Recently, Fudenberg and Karreskog (2020) report evidence highlighting the predictive power of semi-grim strategies in repeated PDs with perfect monitoring, and the behavioral assumption that decisions are made in each round, instead of say once at the start of a session (as in the pure-strategy conjecture), seems intuitive.[6] However, there

[3]Specifically, DF18's observation states that subjects' behavior is described "exactly" even if one "does not allow for any mistakes" .

[4]Specifically, a behavior strategy satisfying $(\sigma_{cc}, \sigma_{cd}, \sigma_{dc}, \sigma_{dd}) = (0.9, 0.3, 0.3, 0.1)$, with varying $\sigma_\emptyset$, is approximately played in all treatments of a data set comprising four experiments.

[5]The only other studies investigating a behavior strategy seem to be Fudenberg et al. (2012), who include the strategy "generous TFT" which randomizes (only) after opponent's defection, and more recently Dvorak and Fehrler (2018). A recent study by Romero and Rosokha (2019) indicates that subjects consider randomizing over single choices even ex-ante.

[6]Here follow the interpretation of behavior in repeated matching pennies games (e.g. Goeree et al., 2003), whereby the description of subjects randomizing say 50-50 each round is simply the best-possible description for the outside observer. Subjects themselves typically do perceive their decisions to be deliberate each round. Similarly, we consider a behavior strategy implying randomization each round to be the best-possible descrip-

are several concerns about Breitmoser's results that might explain why the behavior-strategy conjecture faces skepticism: the data set might be fortunately selected in Breitmoser (2015), behavior might be more complex than memory-1 admits, strategies may be behavior strategies other than semi-grim, subjects might switch strategies as the session progresses, and round-1 behavior was not included in the estimation of strategies. In the next two sections, we address all of these concerns and report arguably conclusive answers to the following questions whose answers then serve as foundation for the structural analysis of preferences and beliefs:

Question 1. *Do subjects play pure, mixed, or behavior strategies?*

Question 2. *Is there heterogeneity in subject types and which strategies are played?*

The case for memory-2 strategies had been made by Fudenberg et al. (2012), who analyze the repeated PD with imperfect monitoring and show that if we assume subjects play pure strategies, then there must be subjects with memory-2, based on evidence for 2TFT and "lenient" Grim2 strategies. Similar ideas are expressed in Aoyagi and Frechette (2009) and Bruttel and Kamecke (2012). Our analysis will allow for memory-2, but in addition, we will relax the restriction to pure strategies, which seems critical since behavior strategies also generate decision patterns resembling memory-2 or -3.

The data We re-analyze the exact same set of experiments reviewed in Dal Bó and Fréchette (2018). This set comprises most of the modern experiments on repeated Prisoner's Dilemmas with perfect monitoring, i.e. those published since Dal Bó (2005), and consists in total of data from 12 experiments, 32 treatments, more than 1900 subjects, and almost 145,000 decisions. The set of experiments equates with the experiments listed in Table 2. A brief review and an overview table is in Appendix B, but for a detailed discussion, see DF18. Due to its enormous size, the wide range of experiments covered (from different experimenters in various universities and various countries), and its comprehensive character with respect to the recent list of experiments on the repeated PD, this data set appears to be optimal for our purposes. In addition, by sticking exactly to the list of experiments reviewed by Dal Bó and Fréchette (2018), we can rule out the notion that data selection biases the results in favor of any of the hypotheses we intend to test.

Econometric approach Our econometric approach is standard, building on finite-mixture and Markov-switching analyses generalizing the strategy frequency estimation method of Dal Bó and Fréchette (2018) as we simultaneously estimate strategies and their frequencies. All details, including a simulation analysis of validity given the finite data sets considered here, are provided in the Appendix, Section A.

tion available to observers of seemingly random but subjectively deliberate (memory-1) decisions that subjects make each round.

3 A model-free overview of behavior

In order to provide a foundation for the subsequent analysis and discussion, let us first provide an overview of behavior in the repeated PD without imposing restrictions reflecting any of the above stated three conjectures. To this end, we simply report average cooperation rates in both the first and second halves of sessions of all experiments and discuss how these average strategies align with expected payoffs across states.

Average behavior Table 2 reports the average cooperation rates across experiments in each of the four memory-1 states after round 1 and tests for significant differences. For brevity, we aggregate across all treatments per experiment here but provide results by treatment in Table 15 in the appendix, then also including round-1 behavior. Initially, we skip round-1 behavior as it varies substantially across treatments, as discussed below, but the cooperation rates in the remaining states are fairly similar across treatments and indeed across experiments, as Table 2 shows. In state cc, cooperation rates are above 0.9, in state dd they are mostly at or below 0.1 (with the sole exception of Aoyagi and Frechette, 2009), and after the mixed histories cd and dc, cooperation rates fluctuate somewhat in the range $[0.2, 0.5]$. Further, the differences between inexperienced and experienced subjects are very minor overall, the aggregate cooperation probabilities shift by at most five percentage points. This observation notwithstanding, it is customary to distinguish experienced and inexperienced behavior, by first and second halves of sessions, which we maintain also for this paper.

Re-analyzing four experiments, Breitmoser (2015) made the observation that average memory-1 strategies have a "semi-grim" pattern. A behavior strategy is called *semi-grim* if $\sigma_{cc} > \sigma_{cd} \approx \sigma_{dc} > \sigma_{dd}$. Based on the vastly extended data set analyzed here, we can scrutinize whether this somewhat surprising observation was the result of a selection effect. We test for differences in the cooperation rates using bootstrapped p-values, resampling at the subject level, and distinguishing two levels of significance: the conventional level 0.05 and the tighter level $0.002 \approx 0.05/24$. The latter implements the Bonferroni correction for tests across 12 experiments and the two session halves. Naturally, we shall focus on this corrected level of significance, but for clarity we also report the conventional level that does not correct for multiple testing.[7]

Out of all the 24 observations, considering first and second halves separately, only one observation, based on one session half in one experiment (Dreber et al., 2008), indicates a significant violation of the key restriction $\sigma_{cd} \approx \sigma_{dc}$, while the other two restrictions $\sigma_{cc} > \sigma_{cd,dc}$ and $\sigma_{cd,dc} > \sigma_{dd}$ are never violated significantly. In $45/48$ cases they are even confirmed significantly at the tight 0.002 level surviving the Bonferroni correction. Pooling all observations from all experiments, $\sigma_{cd} \approx \sigma_{dc}$ is not rejected in the first halves of sessions but at the 0.05 level it is rejected in the second halves of sessions. The difference of σ_{cd} and σ_{dc} remains small, however, and is not significant at the 0.025 level surviving the Bonferroni correction considering that we run two simultaneous tests for the pooled data (one for the first halves of sessions and one for the second halves). Given this range of observations on a

[7] In Table 2, $<, >$ indicate significance at the conventional level and $\ll, \gg$ indicate significance surviving the Bonferroni correction (see the table notes for details).

vastly extended data set, we conclude that Breitmoser's observation passed the out-of-sample test on non-selected data, i.e. that average behavior indeed exhibits the semi-grim pattern.

We want to emphasize that, if there is subject heterogeneity, mean cooperation rates provide unbiased estimates of the true cooperation rates but are not necessarily unbiased estimates of the mean strategies (e.g. due to selection effects after round 1). Yet, the behavior-strategy conjecture postulates that this semi-grim pattern does not only characterize the behavior on average but also the strategies of individual subjects. Otherwise, the observation that this pattern recurs across all treatments and experiments would appear to be a striking coincidence—for, if used at all, pure strategies are estimated to be played in strikingly varying weights across treatments (Dal Bó and Fréchette, 2018), which seems incompatible with the observation that mean cooperation rates always exhibit the semi-grim pattern—but our objective is to test this conjecture directly.

The results of a first simple test of this hypothesis are reported in the last four columns of Table 2. These columns list the number of subjects (per experiment) that deviate significantly from randomizing 50-50 in the four memory-1 states. We focus on subjects with at least five observations per state, which is sufficient to trigger significance in two-sided Fisher tests if subjects play a pure strategy. The results are fairly clear: In state *cd*, i.e. after unilateral defection of the opponent, all standard pure strategies (except AC, which is rarely observed though) agree on the (pure) prediction that one should defect. This state is unique with respect to the unanimity of the prediction. For this state, however, we find the lowest number of subjects significantly deviating from randomizing 50-50—only around a quarter of the subjects do so, putting a rather tight bound on the number of subjects potentially playing pure strategies.

To further illustrate this bound, assume that subjects do use pure strategies: On one hand, given that the semi-grim pattern results on average, there have to be subjects that systematically cooperate after unilateral defection of opponents (state *cd*). These subjects are rarely found in analyses, as indicated most clearly by the aforementioned Result 6 of Dal Bó and Fréchette (2018), stating that "always defect" (AD), Grim, and tit-for-tat (TFT) are the "three strategies [that] account for most of the data". This directly contradicts the observation that $\sigma_{cd} \approx \sigma_{dc} > \sigma_{dd}$, unless in addition to the strategies accounting for most of the data a substantial number of subjects systematically cooperate in state *cd*. However, the strategies predicting at least occasional cooperation after *cd*, such as always-cooperate and tit-for-2-tats, were found to fit behavior of only very few subjects in Dal Bó and Fréchette (2018). This contradiction foreshadows what we will find below: even allowing for drastic data mining, pure strategies cannot be pushed to fit behavior as well as a simple behavior strategy does.

Relation to monetary incentives Complementing the model-free description of behavior, let us look at what subjects should be doing under rational expectations. While relating the decisions "cooperate" and "defect" to expected payoffs in each state is a standard behavioral piece of information in analyses of static games, it is novel in analyses of repeated games. The underlying question, whether the actions chosen are at least qualitatively plausible, is of obvious relevance in any attempt to understand behavior.

Table 2: Few subjects play pure strategies and assuming pure strategies yields a striking bias even in large mixture models

	Actual cooperation rates							Number of subjects not randomizing 50-50			
Experiment	$\hat{\sigma}_{cc}$		$\hat{\sigma}_{cd}$		$\hat{\sigma}_{dc}$		$\hat{\sigma}_{dd}$	(c,c)	(c,d)	(d,c)	(d,d)
First halves per session											
Aoyagi and Frechette (2009)	0.917	$\gg$	0.45	$\approx$	0.408	$\approx$	0.336	32/38	1/23	3/20	7/21
Blonski et al. (2011)	0.89	$\gg$	0.279	$\approx$	0.193	$\gg$	0.034	13/17	1/5	3/3	124/135
Bruttel and Kamecke (2012)	0.91	$\gg$	0.286	$\approx$	0.228	$\gg$	0.08	12/18	6/23	8/21	32/36
Dal Bó (2005)	0.922	$\gg$	0.212	$<$	0.342	$\gg$	0.089	13/13	0/3	2/2	42/54
Dal Bó and Fréchette (2011)	0.951	$\gg$	0.334	$\approx$	0.331	$\gg$	0.063	94/106	28/117	51/128	218/253
Dal Bó and Fréchette (2015)	0.94	$\gg$	0.297	$\approx$	0.335	$\gg$	0.057	216/243	37/137	62/147	404/474
Dreber et al. (2008)	0.904	$\gg$	0.217	$\approx$	0.213	$\gg$	0.036	15/25	3/19	12/18	45/48
Duffy and Ochs (2009)	0.904	$\gg$	0.301	$\approx$	0.33	$\gg$	0.111	43/57	4/25	10/24	61/82
Fréchette and Yuksel (2017)	0.943	$\gg$	0.141	$\approx$	0.266	$\approx$	0.091	21/28	0/0	2/2	5/8
Fudenberg et al. (2012)	0.982	$\gg$	0.4	$\approx$	0.427	$\gg$	0.066	38/43	1/6	5/11	20/25
Kagel and Schley (2013)	0.935	$\gg$	0.263	$\approx$	0.295	$\gg$	0.051	71/81	20/71	32/60	98/111
Sherstyuk et al. (2013)	0.945	$\gg$	0.328	$\approx$	0.371	$\gg$	0.117	37/44	10/36	12/34	41/52
Pooled	0.938	$\gg$	0.304	$\approx$	0.322	$\gg$	0.065	605/713	111/465	202/470	1097/1299
Second halves per session											
Aoyagi and Frechette (2009)	0.958	$\gg$	0.398	$\approx$	0.517	$\approx$	0.375	33/37	0/12	1/12	5/9
Blonski et al. (2011)	0.923	$\gg$	0.287	$\approx$	0.231	$\gg$	0.02	26/32	10/25	11/16	172/178
Bruttel and Kamecke (2012)	0.947	$\gg$	0.221	$\approx$	0.297	$\gg$	0.041	13/15	8/17	9/12	31/35
Dal Bó (2005)	0.92	$\gg$	0.242	$<$	0.388	$\gg$	0.064	18/27	0/3	0/1	50/65
Dal Bó and Fréchette (2011)	0.979	$\gg$	0.376	$\approx$	0.362	$\gg$	0.041	132/137	34/89	62/100	196/215
Dal Bó and Fréchette (2015)	0.976	$\gg$	0.315	$<$	0.402	$\gg$	0.035	340/365	52/162	77/146	448/497
Dreber et al. (2008)	0.917	$\gg$	0.128	$\ll$	0.39	$\gg$	0.009	14/18	6/11	6/12	41/43
Duffy and Ochs (2009)	0.977	$\gg$	0.367	$\approx$	0.391	$\gg$	0.082	80/87	5/35	16/43	60/68
Fréchette and Yuksel (2017)	0.97	$\gg$	0.233	$\approx$	0.398	$\gg$	0.069	33/37	1/6	2/10	20/25
Fudenberg et al. (2012)	0.971	$\gg$	0.487	$\approx$	0.412	$\gg$	0.083	41/44	2/8	4/10	14/17
Kagel and Schley (2013)	0.966	$\gg$	0.262	$\approx$	0.332	$\gg$	0.025	87/90	16/56	30/46	91/97
Sherstyuk et al. (2013)	0.973	$\gg$	0.482	$\approx$	0.437	$\gg$	0.078	44/48	7/24	17/23	23/29
Pooled	0.971	$\gg$	0.327	$<$	0.376	$\gg$	0.039	861/937	141/448	235/431	1151/1278

Note: The "actual cooperation rates" are the relative frequencies estimated directly from the data. The relation signs encode bootstrapped p-values (resampling at the subject level with 10,000 repetitions) where $<,>$ indicate rejection of the Null of equality at $p < .05$ and $\ll,\gg$ indicating $p < .002$. Following Wright (1992), we accommodate for the multiplicity of comparisons within data sets by adjusting p-values using the Holm-Bonferroni method (Holm, 1979). As a result, if a data set is considered in isolation, the .05-level indicated by "$>,<$" is appropriate. If all 24 treatments are considered simultaneously, the corresponding Bonferroni correction requires to further reduce the threshold to $.002 \approx .05/24$, which corresponds with "$\gg,\ll$". Note that all econometric details here exactly replicate Breitmoser (2015), i.e. the statistical tests are not adapted post-hoc. The "number of subjects not randomizing 50-50" indicates the number of subjects with cooperation rates in the various states differing significantly from 50-50 (in subject-level two-sided binomial tests), conditioning on subjects having moved at least five times in the respective state. The required level of significance is set at $p = 0.0625$ such that five observations are sufficient to trigger statistical significance if the subject plays a pure strategy.

For this initial model-free exposition, we will estimate the expected payoffs of cooperate and defect, in each state, from the perspective of an agent who assumes continuation play follows the average relative frequencies of cooperation observed above. These relative frequencies are denoted as the behavior strategy $\sigma = (\sigma_\emptyset, \sigma_{cc}, \sigma_{cd}, \sigma_{dc}, \sigma_{dd})$. Given σ, the expected payoff in state $\omega \in \{\emptyset, cc, cd, dc, dd\}$ is denoted as π_ω, with

$$\pi_\omega = \sigma_\omega \pi_\omega(c) + (1 - \sigma_\omega)\pi_\omega(d), \tag{1}$$

where $\pi_\omega(c)$ and $\pi_\omega(d)$ denote the expected payoffs of playing c and d in state ω,

$$\pi_\omega(c) = \sigma_{\omega'}\big(\delta\pi_{cc} + (1-\delta)\times 1\big) + (1-\sigma_{\omega'})\big(\delta\pi_{cd} + (1-\delta)\times(-l)\big), \tag{2}$$

$$\pi_\omega(d) = \sigma_{\omega'}\big(\delta\pi_{dc} + (1-\delta)\times(1+g)\big) + (1-\sigma_{\omega'})\big(\delta\pi_{dd} + (1-\delta)\times 0\big), \tag{3}$$

with continuation probability δ and ω' the state ω from the opponent's point of view, such that $\sigma_{\omega'}$ is the probability of cooperation by the opponent. By inserting the treatment-specific average behavior strategies σ from above, we can solve the linear equation system, Eqs. 1–3 for all ω, and obtain the expected payoffs $\pi_\omega(c)$ and $\pi_\omega(d)$.

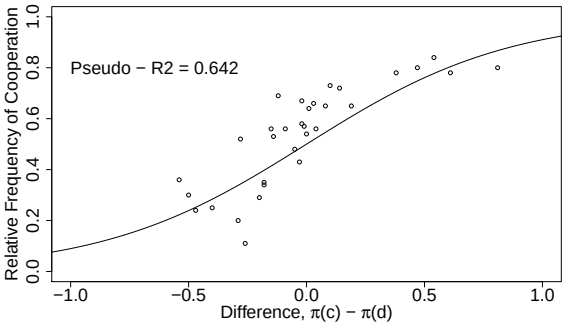
The monetary incentive to cooperate is $\pi_\omega(c) - \pi_\omega(d)$, for each ω. Figure 2 provides an overview of the results: We plot the relative frequencies of cooperation across treatments against the respective monetary incentives to cooperate for each state, separately for first and second halves of sessions. The states cd and dc are pooled for simplicity. Figure 2 additionally shows the best-fitting logistic curve, estimated without intercept such that neutral incentives $\pi_\omega(c) - \pi_\omega(d) = 0$ yield a predicted cooperation probability of 0.50. The pseudo-R^2 of the logistic curves indicate how much of the null deviance is explained by allowing for logistic errors in utility maximization.

The observations can be summarized as follows: For each state, we have observations from treatments with net incentives ranging from around -0.5 to $+1$, i.e. from cases where $\pi_\omega(c) - \pi_\omega(d)$ is highly negative to cases where it is highly positive. Essentially, the former obtains in treatments where Grim is not a subgame-perfect equilibrium strategy and the latter obtains in treatments where the discount factor δ is substantially above the threshold for Grim to be a subgame-perfect equilibrium strategy. Despite this range of induced monetary incentives, relative probabilities of cooperation and monetary incentives are highly correlated only in round 1 (state $\emptyset$). They are statistically close to independent in all states after round 1. For example, in second halves of sessions, when subjects have gained experience, the Pseudo-R^2 of the logit model is above 0.8 in round 1 and below 0.2 in all states afterwards. Obviously, this model-free analysis has the drawback of neglecting subject heterogeneity, which we will address below, but it seems that behavior in states cc and dd may be difficult to align with monetary incentives. For this reason, we raise the following set of questions.

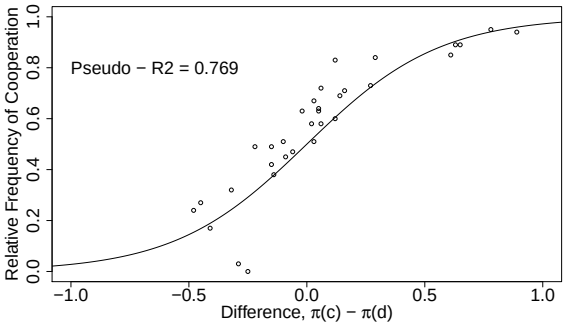
Question 3. *Do subjects act rationally and with rational expectations in round 1 but irrationally follow some automaton or heuristic afterwards? How do strategies relate to treatment parameters? Can we rationalize choices after round 1? And are the strategies predictable?*

Figure 2: Relation of monetary incentives and cooperation rates across states (naive beliefs)

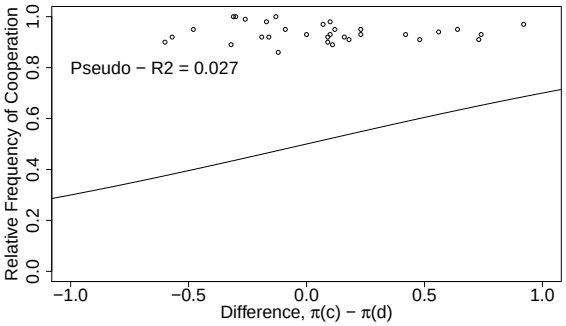
(a) State $\emptyset$, first halves of session



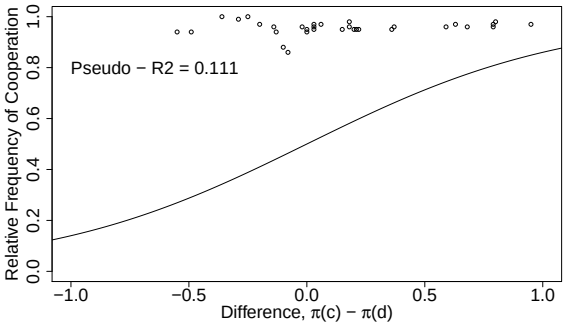
(b) State $\emptyset$, second halves of session



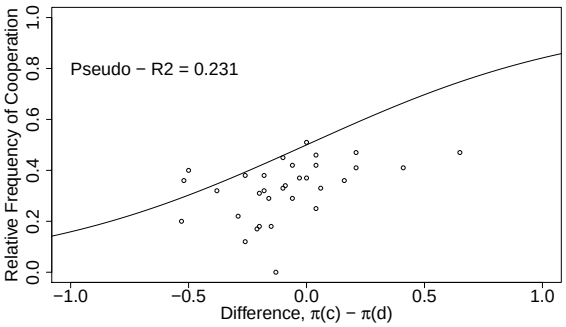
(c) State cc, first halves of session



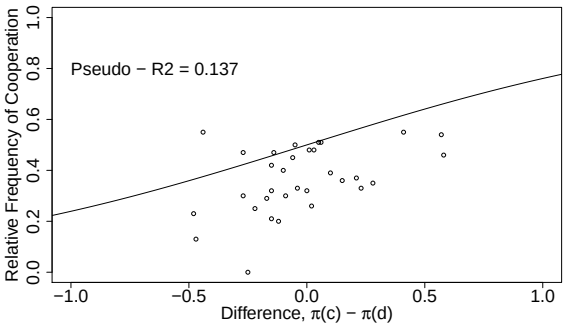
(d) State cc, second halves of session



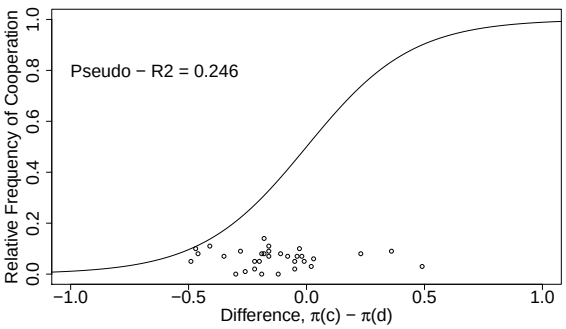
(e) State cd, dc, first halves of session



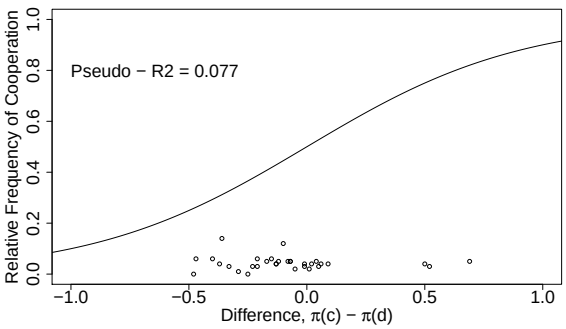
(f) State cd, dc, second halves of session



(g) State dd, first halves of session



(h) State dd, second halves of session



Note: For further information, set Tables 52–59 in the supplement.

4 Estimating the strategies used by subjects

This section consists of two parts. In the first part, we data mine for the best possible (post-hoc) mixtures of (generalized) pure strategies for *each treatment*. We will not penalize the model for data mining best mixtures but treat the resulting mixtures treatment-by-treatment as if they had been hypothesized ex-ante. As we discuss below, this provides us with an upper bound for the goodness-of-fit of pure and mixed strategies, which we will compare to a simple model that contains only defectors playing AD and cooperators playing semi-grim behavior strategies (as previously defined in Breitmoser, 2015) in all treatments. Due to the one-sided data mining, this analysis is heavily lopsided in favor of modeling behavior using pure and mixed strategies, and in this sense, we give the pure- and mixed-strategy conjectures the best possible chance.

In the second part, we provide the results of an unrestricted estimation of memory-1 strategies, and then estimate the number of subject types and the strategies played in both a top-down and a bottom-up approach towards model selection. The top-down approach starts with the general model and iteratively eliminates insignificant components, while the bottom-up approach starts with a basic model and iteratively adds model components identified as significant.

Both parts of this section will converge to the same model distinguishing defectors playing AD from cautious and strong cooperators playing semi-grim strategies. Their behavior will be further analyzed in the next section. Section B in the appendix demonstrates robustness to longer memory lengths by demonstrating that model adequacy does not improve by equipping subjects with memory-2, neither for (generalizations of) pure strategies nor for semi-grim. That is, while increasing memory length slightly improves the goodness-of-fit, this increase does not make up for the increased complexity of strategies as evaluated using the Bayesian information criterion.

Pure, mixed or behavior strategies? In order to outline our approach towards estimating an upper bound of the goodness-of-fit of pure strategies, recall that the pure memory-1 strategies AD, TFT, and Grim had been conjectured to capture the behavior of most subjects across treatments. For reasons discussed shortly, we add AC and WSLS to obtain a set of baseline strategies. We then extend this set of strategies in two ways. On one hand, we add generalized versions of these strategies by introducing a free parameter per strategy to relax assumptions on first-round cooperation rates $\sigma_\emptyset$, thus allowing subjects' first-round cooperation rates to be different from 0 in AD, and different from 1 in all other strategies. The definition of the continuation behavior remains unchanged, such that $(\sigma_{cc}, \sigma_{cd}, \sigma_{dc}, \sigma_{dd}) \in \{0,1\}^4$ for all pure strategies. We refer to these strategies as *generalized pure strategies of type I*. On the other hand, in the set of *generalized pure strategies of type II*, we introduce a free parameter to allow for randomization within supergames to relax assumptions about behavior after histories such as *cd* or *dc*, where the pure strategies tend to fit poorly. Using the notation introduced above, defining strategies as quintuple $(\sigma_\emptyset, \sigma_{cc}, \sigma_{cd}, \sigma_{dc}, \sigma_{dd})$, generalized TFT is defined as $(1,1,0,\theta^{TFT},0)$, generalized Grim as $(1,1,\theta^G,\theta^G,\theta^G)$, and generalized WSLS as $(1,1,0,0,\theta^{WSLS})$. Generalized AC and AD are defined as behavior strategies

$(1, \theta^{AC}, \theta^{AC}, 0, 0)$ and $(0, \theta^{AD}, \theta^{AD}, \theta^{AD}, \theta^{AD})$, respectively, with all $\theta^* \in [0,1]$.[8] The advantage of defining generalized strategies this way is that linear combinations of these generalized strategies, or of the original pure strategies, can reproduce the aggregate semi-grim patterns we observed above. In addition, we will of course consider the pure strategies in their original form, thereby covering the possibility that in at least some treatments neither of the generalizations improves the goodness-of-fit, allowing us to post-hoc save parameters. In addition to all of this, we allow for trembling-hand noise, i.e. that subjects may deviate from the assumed (generalized) pure strategy with probability $\varepsilon \in [0,1]$ in any given round, to then randomize uniformly.

With this set of strategies in hand, our approach toward data mining the mixtures across treatments is as follows.

First, we evaluate independently for each treatment which mixture of pure or generalized pure strategies best captures behavior. That is, we determine for each treatment, which combination of *pure* strategies fits best, which combination of *generalized pure strategies of type I* fits best, which of *type II*, and which of the three best combinations fits best. Following the pure-strategy conjecture, we assume the best combination always contains at least TFT, AD, and Grim. We add the remaining strategies when this improves the goodness-of-fit. Thus, we choose the best out of 13 as promising conjectured memory-1 mixtures, for each of the 32 treatments and each of the two half-sessions independently.[9] In total, we therefore evaluate 13^{32} models per level of experience and afterwards pick the best-fitting model in terms of ICL-BIC (see Appendix A). Second, we do all of this separately for the three "switching models" designed to capture the three possibilities of strategy switching between supergames: "No Switching" (pure strategy), "Random Switching" (mixed strategy), and "Markov Switching" (strategy switching between supergames follows a Markov process), see Appendix A.1 for details.

The results for each of the three switching models are reported in columns 2-5 of Table 3. The leftmost column contains the results for the baseline model comprising AC, AD, TFT, Grim, and WSLS without data mining, which can serve as a reference for how much of the goodness-of-fit is due to data mining. For the sake of readability, we report ICL-BICs aggregated by experiment.[10]

The random switching model in column 3 of Table 3 capturing mixed strategies generally fits worst, by the enormous amount of more than 2000 points on the log-likelihood scale.

[8] Allowing for more than one free parameter per generalized pure strategy would be unreasonable since they would not be similar enough to their name giving pure strategy anymore. In addition, the penalty for free parameters would increase strongly.

[9] For each of the three classes of strategies (pure, generalized type I, generalized type II), we consider mixtures containing AD, TFT and Grim and in addition either (i) no other strategy, (ii) AD, (iii) WSLS, and (iv) AD + WSLS. This makes 12 combinations in total. In addition, in the case of pure strategies, we allow for a mixture containing noise players (randomizing 50-50 in all states) as type besides AC, TFT and Grim, which are otherwise contained as special case in generalized strategies of type II.

[10] Treatment-wise ICL-BICs are provided in the Appendix E, after Table 21. Each entry in the aggregated table represents the sum of ICL-BICs of the best out of 13 models for each respective treatment. Tables 20 and 21 in the appendix contain additional details regarding the intermediate results obtained during data mining for the best model.

This shows that subjects are reasonably consistent in their strategy choice. The no-switching model capturing pure strategies (column 2) fits worse than the Markov-switching model (column 4) in the first halves of sessions, but weakly better in the second halves of sessions. If these models captured behavior well, this could suggest that subjects initially experiment with different pure strategies, though not randomly, as in mixed strategies, but systematically, as in a stochastic Markov process, to then converge to individual choices for strategies as the session progresses. Additionally, Table 20 (in the appendix) shows that *continuation strategies* of the generalized pure type II (excluding round-1 behavior) perform much better than their counterparts without generalization. The differences in model fit are large, amounting in total to more than 1000 points on the log-likelihood scale, which suggests that randomization within supergames is indeed a behavioral facet.

However, the arguably most relevant observation at this point concerns the aggregate effect achieved by data mining for the best-fitting combination of pure strategies and switching model. Modeling the behavior of inexperienced subjects (first halves of sessions), our generalizations and data mining combined yield a gain of 2000 points on the log-likelihood scale, comparing the baseline model to the best-fitting Markov switching models, and modeling experienced subjects (second halves), generalization and data mining combined yield a gain of 2500 points compared to the baseline model. Since these scores do not account for the degrees of freedom inherent in the model selection during data mining, they do not imply that the baseline model has to be rejected, but they clearly show that our approach yields an enormous improvement in fit over the standard memory-1 mixtures typically proposed in the literature. Further, since we attempted to include all specifications that may be considered compatible with either the pure- or the mixed-strategy conjecture, and picked the best one for each treatment, we can consider this data-mined specification to be a generous upper bound of the adequacy of these memory-1 models to describe behavior.

Second, this upper bound, reported in column 5 ("Best Switching") of Table 3, allows us to test the pure- and mixed-strategy conjectures against the behavior-strategy conjecture. Since the behavior-strategy conjecture regards the behavior of cooperating subjects, we complete the model by allowing for AD players next to semi-grim players to capture the behavior of defecting subjects (in a no-switching model).

Note that we compare this simple and constant two-type model defined prior to the analysis, with 5 free parameters per treatment,[11] to the "Best Switching" model that was post-hoc picked from 3×13^{32} models, after estimating 438 parameters for each of the 32 treatments but without accounting for the degrees of freedom used in the model selection process (solely accounting for the 3–10 parameters of the best-fitting model that is finally used in line with the data mining ideal). The results are reported in column 6 ("AD + SG"). Despite this abuse of statistical power, the simple AD+SG model fits significantly better than the mined mixture of generalized pure or mixed strategies: it improves on the data-mined model by another 400 points in the first-halves of sessions and even by 750 points in the second halves of sessions. Since AD players are contained in all models, this demonstrates that the behavior of subjects not playing AD—i.e. behavior of cooperating subjects—is much

[11] Three parameters for the semi-grim strategy $(\theta_1^{SG}, \theta_2^{SG}, \theta_3^{SG}, \theta_3^{SG}, 1-\theta_2^{SG})$, one for the share of AD players, and one to capture noise of the AD players.

Table 3: **Best mixtures of pure or generalized strategies in relation to semi-grim.** Strategy mixtures are estimated treatment-by-treatment. The resulting ICL-BICs are pooled for experiments and overall (less is better, relation signs point to better models)

		Best mixture of pure or generalized strategies						Best Mixture
	Baseline Model	No Switching	Random Switching	Markov Switching	Best Switching	AD + SG	AD + 2SG	Best Switching By Treatment
Specification								
# Models evaluated	1	13^{32}	13^{32}	13^{32}	3×13^{32}	1	1	$39^{32} \approx 10^{51}$
# Pars estimated (by treatment)	5	80	80	278	438	5	9	438
# Parameters accounted for	5	3–10	3–10	12-35	3–30	5	9	3–30
First halves per session								
Aoyagi and Frechette (2009)	886.44 $\gg$	756.95 $\approx$	763.11 $\approx$	755.97	755.97 $\approx$	781.86 $\approx$	777.3 $\approx$	755.97
Blonski et al. (2011)	1114.69 $\gg$	1069.58 $\approx$	1104.85 $\ll$	1225.35	1225.35 $\gg$	1069.28 $<$	1134.96 $>$	1069.39
Bruttel and Kamecke (2012)	845.41 $\approx$	817.89 $\approx$	835.6 $>$	785.49	785.49 $\approx$	800.12 $>$	762.83 $\approx$	785.49
Dal Bó (2005)	666.1 $>$	635.04 $<$	674.57 $\approx$	648.75	648.75 $\approx$	629.17 $\approx$	600.26 $<$	631.2
Dal Bó and Fréchette (2011)	7423.23 $\gg$	6904.79 $\ll$	7456.12 $\gg$	6388.49	6388.49 $<$	6597.93 $\gg$	6304.97 $\approx$	6388.49
Dal Bó and Fréchette (2015)	8880.62 $\gg$	8434.93 $\ll$	9166.72 $\gg$	8158.31	8158.31 $>$	8017.59 $\gg$	7810.7 $\ll$	8138.61
Dreber et al. (2008)	871.32 $\gg$	787.71 $<$	863.7 $\gg$	752.16	752.16 $\approx$	782.37 $\approx$	763.52 $\approx$	752.16
Duffy and Ochs (2009)	1448.71 $\approx$	1395.4 $<$	1461.01 $>$	1372.99	1372.99 $\approx$	1372.97 $\approx$	1320.71 $\approx$	1372.99
Fréchette and Yuksel (2017)	321.32 $\approx$	300.87 $<$	337.5 $>$	298.53	298.53 $\approx$	299.62 $\approx$	284.11 $\approx$	298.53
Fudenberg et al. (2012)	454.09 $\approx$	432.32 $\approx$	432.38 $\approx$	425.54	425.54 $>$	381.01 $\approx$	370.01 $<$	425.54
Kagel and Schley (2013)	2735.02 $\approx$	2685.4 $\ll$	2993.4 $\gg$	2439.06	2439.06 $\approx$	2561.76 $\gg$	2421.27 $\approx$	2439.06
Sherstyuk et al. (2013)	1389.33 $\approx$	1322.6 $\ll$	1450 $\gg$	1296.85	1296.85 $\approx$	1303.8 $\gg$	1200.28 $<$	1296.85
Pooled	**27218.66** $\gg$	**25758.38** $\ll$	**27754.81** $\gg$	**25166.24**	**25166.24** $>$	**24779.85** $\gg$	**24079.18** $\ll$	**24863.15**
Second halves per session								
Aoyagi and Frechette (2009)	534.29 $\gg$	416.51 $\approx$	437.8 $\approx$	423.05	416.51 $\approx$	423.68 $\approx$	422.24 $\approx$	416.51
Blonski et al. (2011)	1503.41 $\gg$	1398.5 $\ll$	1509.09 $<$	1593.01	1398.5 $>$	1346.79 $\approx$	1385.91 $\approx$	1394.16
Bruttel and Kamecke (2012)	588.33 $>$	538.17 $<$	611.91 $\gg$	516.71	538.17 $\approx$	536.77 $>$	478.23 $\approx$	516.71
Dal Bó (2005)	751.82 $\approx$	732.27 $<$	786.21 $>$	739.59	732.27 $>$	699.05 $\approx$	679.04 $<$	729.48
Dal Bó and Fréchette (2011)	6065.93 $\gg$	5195.88 $\ll$	6378.16 $\gg$	5007.24	5195.88 $\approx$	5128.69 $\gg$	4545.08 $\ll$	4964.77
Dal Bó and Fréchette (2015)	9085.4 $\gg$	8177.46 $\ll$	9401.19 $\gg$	7910.83	8177.46 $\gg$	7825.98 $\gg$	7310.27 $\ll$	7893.79
Dreber et al. (2008)	656.38 $\approx$	619.9 $\approx$	662.24 $>$	581.94	619.9 $\approx$	589.84 $>$	541.83 $\approx$	581.94
Duffy and Ochs (2009)	2010.01 $>$	1883.52 $\approx$	1914.83 $>$	1850.35	1883.52 $>$	1761.6 $>$	1602.93 $\ll$	1850.35
Fréchette and Yuksel (2017)	469.85 $\approx$	433.18 $<$	474.93 $>$	427.79	433.18 $\approx$	423.34 $\approx$	381.63 $\ll$	427.79
Fudenberg et al. (2012)	530.3 $\approx$	514.87 $\approx$	516.12 $\approx$	515.97	514.87 $\approx$	452.6 $>$	414.24 $<$	514.87
Kagel and Schley (2013)	1866.19 $\approx$	1751.81 $\ll$	2336.29 $\gg$	1678.7	1751.81 $\approx$	1775.62 $\gg$	1541.38 $<$	1678.7
Sherstyuk et al. (2013)	1027.43 $>$	955.73 $\ll$	1137.49 $\gg$	958.99	955.73 $\approx$	951.34 $\gg$	823.06 $\ll$	955.73
Pooled	**25271.72** $\gg$	**22848.49** $\ll$	**26409.44** $\gg$	**22927.9**	**22848.49** $\gg$	**22097.67** $\gg$	**20454.13** $\ll$	**22422.07**

Note: Results treatment-by-treatment are in the appendix. Relation signs encode p-values of Schennach-Wilhelm likelihood-ratio tests where $<, >$ indicate rejection of the Null of equality at $p < .05$ and $\ll, \gg$ indicating $p < .002$, which implements the Bonferroni correction of 24 simultaneous tests per hypothesis. "No Switching" assumes that subjects chooses a strategy prior to the first supergame and plays this strategy constantly for the entire half session. "Random Switching" assumes that subjects randomly chooses a strategy prior to each supergame (by i.i.d. draws), and "Markov Switching" allows that these switches follow a Markov process.

better described by the semi-grim behavior strategy than using any mixture of received or generalized pure strategies. This is substantial and perhaps surprising, but in the end, it is simply a reflection of the deficiency of deterministic choice rules in capturing behavior discussed above. A robustness check clarifying that this observation also holds true after accounting for memory-2 is reported in the appendix.

Third, the AD+SG model provides an estimate of the lower bound of the goodness-of-fit that can be achieved when we give up the restriction to pure strategies and allow subjects to play behavior strategies. To provide a first indication of the heterogeneity of cooperating subjects, we also report the results for a model ("AD + 2SG" in column 7) that allows for two different cooperating subject types, each playing a semi-grim strategy, alongside AD players. This specification turns out to improve goodness-of-fit significantly for both experience levels, by another 700 points in first halves and even 1600 points in second halves of sessions. The heterogeneity of cooperating subjects will be investigated below in the unrestricted analysis.

Fourth, we evaluate the arguably extreme model, which identifies the best-fitting combination of (generalized) pure strategies (out of 13 combinations) *and* the best-fitting switching model (out of 3) treatment by treatment without any consistency requirement. Thus, we choose the best-fitting model from 39 models for each treatment, amounting to the enormous selection of the best out of 39^{32} models across all experiments. Note that such analysis without imposing consistency requirements across treatments does not yield economically useful estimates, but if anything, this provides an even more generous upper bound on the economic content of pure and generalized pure strategies of memory-1. The results are reported in the right-most column ("Best Switching By Treatment"). In total, this exhaustively mined model still fits weakly worse than the AD+SG model predicted by the behavior-strategy conjecture and it fits highly significantly worse (by more than 2000 log-likelihood points for experienced subjects) than the "AD+2SG" model that allows for heterogeneity of cooperating subjects.[12] We summarize these observations as follows.

Result 1 (Question 1)**.** *Cooperating subjects seem to use memory-1 behavior strategies. The upper bound of behavior that can be captured with received pure or mixed strategies is significantly lower than the adequacy of a model assuming all cooperating subjects play the same (semi-grim) behavior strategy.*

Heterogeneity of cooperators and unrestricted estimation Let us now examine to what extent the cooperating subjects are heterogeneous and indeed play semi-grim strategies. In order to test this joint hypothesis of heterogeneity and semi-grim, let us start with a general model allowing for four different subject types (per treatment), one of which plays AD and three that play general memory-1 behavior strategies without imposing restrictions such as semi-grim.[13] In Table 4, we refer to this model as "$3 \times P5 + AD$", where $P5$ indicates use of an unrestricted five-parameter behavior strategy. Table 4 provides detailed information

[12] Section B in the appendix demonstrates that this result is robust to allowing for memory-2, where we find that memory-2 is overall insignificant.

[13] As a reminder, these restrictions are $\sigma_{cd} = \sigma_{cd}$ and $\sigma_{cc} = 1 - \sigma_{dd}$.

on a range of models that distinguish either up to three cooperating types playing general behavior strategies or up to three types playing semi-grim strategies. This will allow us to directly test the joint hypothesis.

Before doing so, let us point to an arguably important observation. Table 4 reports on a large range of models where cooperating subjects always are assumed to play behavior strategies. All of these models improve on the best of the 10^{51} models assuming subjects play pure or generalized pure strategies ("Best Mixture, Best Switching" in the left-most column of Table 4). That is, our earlier result on the inadequacy of pure and generalized pure strategies is confirmed very robustly: whatever specification we use, allowing cooperating subjects to play behavior strategies fits behavior much better. That is, the best of the 10^{51} models assuming pure or generalized pure strategies fits at least weakly worse than the worst of the seven models assigning cooperating subjects behavior strategies, and significantly worse than all of the five models allowing for at least two distinct types of cooperating subjects. Notably, this would not be observed if the pure-strategy conjecture was empirically valid: The unrestricted analysis allows cooperating subjects to play, besides AD, the cooperative strategies TFT, Grim and say WSLS or AC depending on treatment (in $3 \times P5 + AD$), and if they actually did so, then the pure strategy mixture would fit substantially better without using as many free parameters.

Now, using "$3 \times P5 + AD$" as the baseline, we can analyze which form of heterogeneity is most suitable for describing behavior. Starting with four subject types seems to be sufficient ex-ante, and will turn out to be sufficient ex-post. In Table 4, the two right-most columns report on the adequacy of nested models that distinguish only two types or one type of cooperating subjects (besides the AD type). It turns out that distinguishing just two types of cooperating subjects ("$2 \times P5 + AD$") weakly improves on distinguishing three types, while models with just one cooperating type ("$P5 + AD$") fit significantly worse. The latter corresponds with our finding above (prior to Result 1) that cooperating subjects are not homogeneous.

To the left of column "$3 \times P5 + AD$", Table 4 details information on models assuming the cooperating subjects play semi-grim strategies rather than unrestricted memory-1 strategies. To be exhaustive, we consider models distinguishing three semi-grim types ("$3 \times SG + AD$"), two semi-grim types ("$2 \times SG + AD$"), one semi-grim type ("$SG + AD$") and 1.5 semi-grim types ("$1.5 \times SG + AD$"). Slightly abusing notation, 1.5 semi-grim types indicates that the two cooperating types have different cooperation probabilities in round 1 of each supergame but equivalent continuation strategies. At this point, the discussion can be kept rather short as the results are fairly clear: All models distinguishing at least two types of cooperating subjects, regardless of which strategy they are assumed to be playing, fit about equally well. The differences between these models are at best weakly significant, while all of them fit significantly better than the two models assuming cooperating subjects are homogeneous ("$SG + AD$" and "$P5 + AD$").

Thus, we find strong evidence for heterogeneity and the behavior-strategy conjecture, though we need additional guidance for or against modeling the behavior strategy as semi-grim. For additional guidance, we can rely on either the top-down or the bottom-up approach towards model selection. By the top-down approach, we start with the most general model

Table 4: Examining heterogeneity of cooperating subjects and semi-grim structure of their strategies

	Best Mixture Best Switching		SG + AD		1.5×SG+AD		2×SG+AD		3×SG+AD		3×P5+AD		2×P5+AD		P5+AD
Specification															
# Models evaluated	$39^{32} \approx 10^{51}$		1		1		1		1		1		1		1
# Pars estimated (by treatment)	438		5		7		9		13		19		17		11
# Parameters accounted for	3-30		5		7		9		13		19		17		11
First halves per session															
Aoyagi and Frechette (2009)	755.97	$\approx$	781.86	$\approx$	792.51	$\approx$	777.3	$\approx$	782.13	$>$	741.95	$\approx$	744.86	$\approx$	744.06
Blonski et al. (2011)	1069.39	$\approx$	1069.28	$\approx$	1104.6	$\approx$	1134.96	$\ll$	1232.97	$\ll$	1332.48	$\gg$	1205.47	$\gg$	1106.01
Bruttel and Kamecke (2012)	785.49	$\approx$	800.12	$\approx$	771.14	$\approx$	762.83	$\approx$	748.06	$\approx$	751.86	$\approx$	759.45	$\approx$	803.58
Dal Bó (2005)	631.2	$\approx$	629.17	$\approx$	618.39	$\approx$	600.26	$\ll$	626.56	$\approx$	639.8	$>$	609.1	$\approx$	620.38
Dal Bó and Fréchette (2011)	6388.49	$<$	6597.93	$>$	6352.59	$\approx$	6304.97	$\approx$	6198.12	$\approx$	6216.22	$<$	6295.32	$\ll$	6553.25
Dal Bó and Fréchette (2015)	8138.61	$\approx$	8017.59	$\gg$	7830.12	$\approx$	7810.7	$\approx$	7828.38	$\approx$	7829.74	$\approx$	7775.7	$\ll$	7969.32
Dreber et al. (2008)	752.16	$\approx$	782.37	$\approx$	764.44	$\approx$	763.52	$\approx$	766.77	$\approx$	765.81	$\approx$	767.32	$\approx$	783.45
Duffy and Ochs (2009)	1372.99	$\approx$	1372.97	$\approx$	1361.15	$\approx$	1320.71	$\approx$	1297.84	$\approx$	1291.42	$<$	1345.16	$\approx$	1361.86
Fréchette and Yuksel (2017)	298.53	$\approx$	299.62	$\approx$	289.54	$\approx$	284.11	$\approx$	289.88	$\approx$	294.05	$\approx$	285.33	$\approx$	291.69
Fudenberg et al. (2012)	425.54	$>$	381.01	$\approx$	377.96	$\approx$	370.01	$\approx$	380.86	$\approx$	381.34	$\approx$	372.32	$\approx$	377.33
Kagel and Schley (2013)	2439.06	$\approx$	2561.76	$\gg$	2450.24	$\approx$	2421.27	$\approx$	2385.02	$\approx$	2354.05	$\approx$	2398.74	$\ll$	2551.68
Sherstyuk et al. (2013)	1296.85	$\approx$	1303.8	$\approx$	1234.52	$\approx$	1200.28	$\approx$	1184.82	$\approx$	1177.24	$\approx$	1186.92	$\ll$	1286.14
Pooled	**24863.15**	$\approx$	**24779.85**	$\gg$	**24202.51**	$\approx$	**24079.18**	$\approx$	**24195.57**	$<$	**24468.99**	$>$	**24219.87**	$\ll$	**24704.09**
Second halves per session															
Aoyagi and Frechette (2009)	416.51	$\approx$	423.68	$\approx$	421.21	$\approx$	422.24	$\approx$	423.63	$>$	404.95	$\approx$	408.59	$\approx$	409.04
Blonski et al. (2011)	1394.16	$\approx$	1346.79	$\approx$	1370.16	$\approx$	1385.91	$<$	1442.85	$\ll$	1555.48	$\gg$	1453.1	$\gg$	1379.87
Bruttel and Kamecke (2012)	516.71	$\approx$	536.77	$\gg$	480.47	$\approx$	478.23	$\approx$	470.25	$\approx$	443.83	$\approx$	471.73	$<$	528.54
Dal Bó (2005)	729.48	$>$	699.05	$\approx$	677.24	$\approx$	679.04	$<$	697.21	$\approx$	707.25	$\approx$	687.86	$\approx$	696.41
Dal Bó and Fréchette (2011)	4964.77	$\approx$	5128.69	$\gg$	4565.93	$\approx$	4545.08	$\approx$	4426.48	$\approx$	4461.98	$\approx$	4493.1	$\ll$	5045.34
Dal Bó and Fréchette (2015)	7893.79	$\approx$	7825.98	$\gg$	7306.25	$\approx$	7310.27	$>$	7170.25	$\approx$	7089.56	$\approx$	7151.84	$\ll$	7683.76
Dreber et al. (2008)	581.94	$\approx$	589.84	$>$	544.66	$\approx$	541.83	$\approx$	539.47	$\approx$	519.28	$\approx$	518.82	$<$	562.99
Duffy and Ochs (2009)	1850.35	$\approx$	1761.6	$\gg$	1656.55	$\approx$	1602.93	$>$	1518.65	$\approx$	1509.7	$<$	1598.04	$\ll$	1715.88
Fréchette and Yuksel (2017)	427.79	$\approx$	423.34	$\approx$	422.61	$\approx$	381.63	$\approx$	375.03	$\approx$	384.11	$\approx$	382.16	$<$	409.93
Fudenberg et al. (2012)	514.87	$\approx$	452.6	$\approx$	433.74	$\approx$	414.24	$\approx$	405.22	$\approx$	410.69	$\approx$	421.81	$\approx$	448.37
Kagel and Schley (2013)	1678.7	$\approx$	1775.62	$\gg$	1572.95	$\approx$	1541.38	$>$	1488.49	$\approx$	1477.87	$\approx$	1527.47	$\ll$	1748.01
Sherstyuk et al. (2013)	955.73	$\approx$	951.34	$\gg$	834.74	$\approx$	823.06	$\approx$	799.39	$\approx$	801.53	$\approx$	815.26	$\ll$	935.01
Pooled	**22422.07**	$\approx$	**22097.67**	$\gg$	**20541.83**	$\approx$	**20454.13**	$>$	**20231.09**	$<$	**20459.26**	$\approx$	**20403.95**	$\ll$	**21818.45**

Note: This table verifies a number of possible mixtures involving semi-grim types as a robustness check for the sufficiency of focussing on the mixtures examined above. E.g. "3× SG refers to a model containing three different versions of memory-1 semi-grim with allowing for heterogeneity of randomization parameters across subjects.

$(3 \times P5 + AD)$ and successively reduce its complexity until such reductions dampen its adequacy significantly. The simplest model that we reach this way without a significantly negative impact on adequacy is $1.5 \times SG + AD$. In turn, by the bottom-up approach, we start with the simplest model ($SG + AD$) and successively increase its complexity as long as these increments significantly improve model adequacy. Starting with $SG + AD$, adequacy improves significantly by allowing for subject types differing in their round-1 behavior ($1.5 \times SG + AD$), in both the first and the second halves of sessions, but beyond that, further increments again are not significant in a manner surviving the Bonferroni correction (indicated by $\gg$ or $\ll$ in Table 4).

That is, both the top-down and the bottom-up approach converge to the same conclusion that we need to distinguish two types of cooperating subjects, whose behavior differs only in round 1 of each supergame. On average, the less cooperative type cooperates with probabilities in $[0.2, 0.5]$ in round 1, similar to the cooperation probabilities after mixed histories cd/dc, and the more cooperative type cooperates with probabilities above 0.9 in most treatments, similar to cooperation probabilities after cc. Table 9 in the appendix provides detailed results.

Result 2 (Question 2)**.** *The analysis identifies two types of cooperating subjects playing the same semi-grim continuation strategy but different cooperation probabilities in round 1 (**cautious cooperators** and **bold cooperators**) and a subject type playing a strategy close to always defect (**defectors**).*

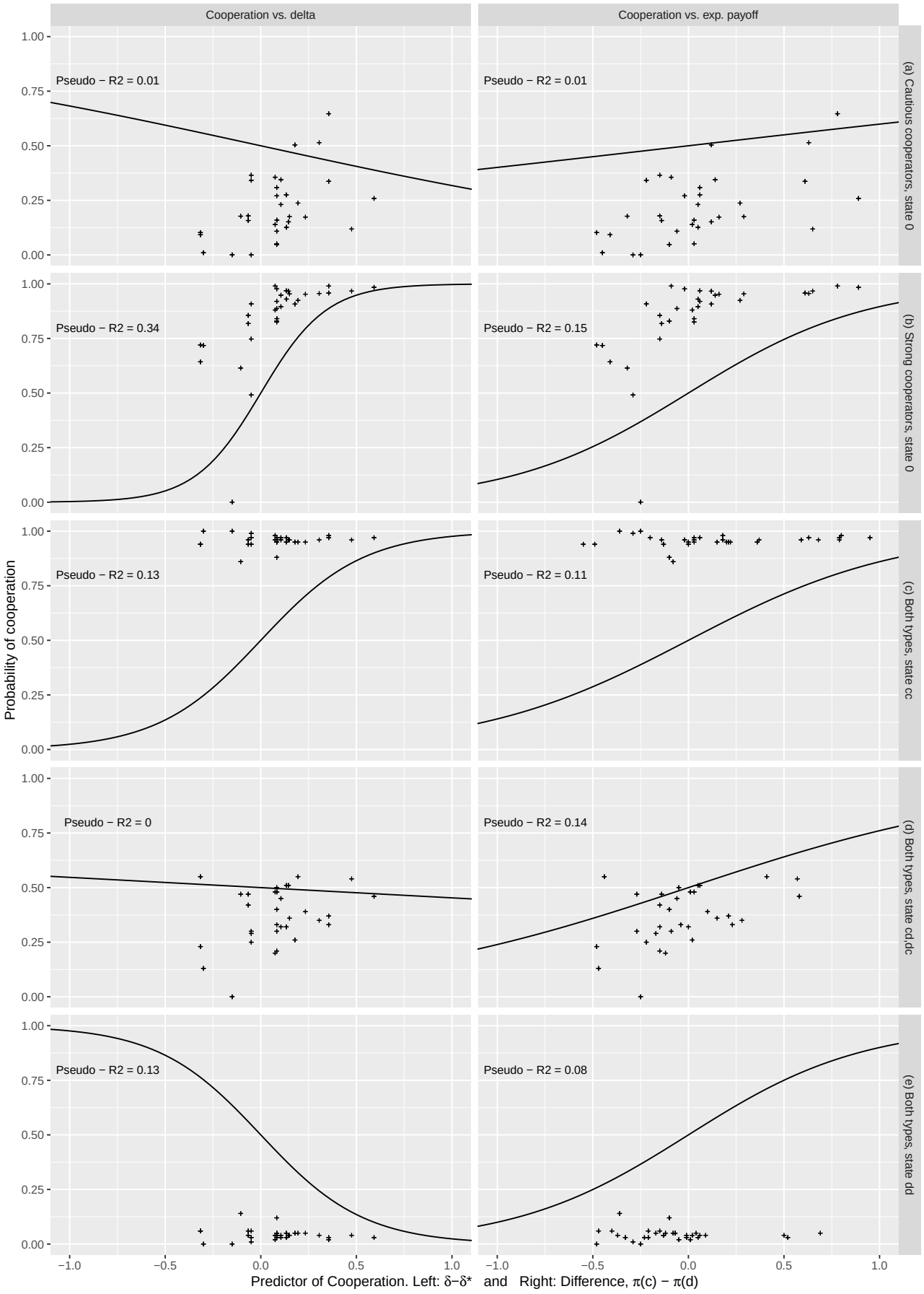
The model with this subject composition, and any other model allowing for two types of cooperating subjects playing behavior strategies, fits significantly better than all 10^{51} models assuming pure or generalized pure strategies.

5 How do strategies relate to supergame parameters?

Having estimated the number of subject types and their strategies, we can revisit Question 3 and ask to what extent the subjects' strategies are functions of treatment parameters, to what extent they are rationalizable, and to what extent they may be predictable. In light of the above results, we distinguish defecting and cooperating subjects. The defecting subjects play slightly perturbed strategies close to AD, which are essentially invariant to treatment parameters and rationalizable to the extent that AD is rationalizable (note that AD is a best response to itself in all supergames considered here). For this reason, we shall focus on the strategies played by cooperating subjects. By Result 2, there are two types of cooperating subjects, both identified as playing semi-grim supergame strategies with significant differences found in the probability of cooperation in round 1.

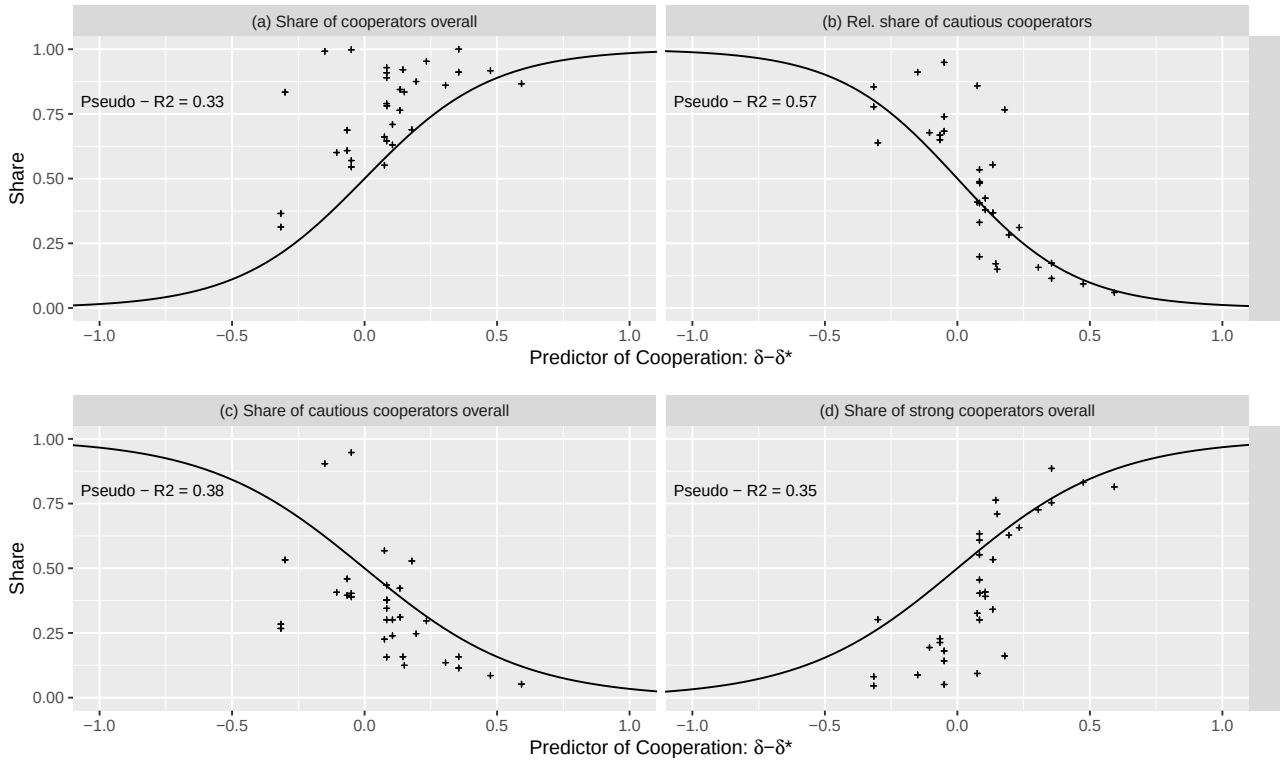
Overview Recall that Figure 2 plotted the average cooperation rates across states against the expected payoffs from cooperation, which suggested that subjects act highly rationally in round 1 but ignore expected payoffs afterwards. We suspected confounds due to looking at raw cooperation rates, most notably possible selection effects, and our estimates of

Figure 3: Relation of δ **(left) and monetary incentives (right)** to cooperation rates (second halves of sessions)



Note: For further information, set Tables 52–59 in the supplement.

Figure 4: Relation of $\delta - \delta^*$ to shares of cooperators (second halves of sessions)



Note: This figure shows how the ratios of the three strategies – defectors, cautious cooperators, and strong cooperators change with the distance of δ to the BOS cooperation threshold δ^* across treatments. The solid line represents the best fitting logistic curve estimated without intercept such that the share is 0.5 for $\delta = \delta^*$. Panel (a) displays the total share of both cooperators, panel (b) the relative share of cautious cooperators among cooperators, panel (c) the share of cautious cooperators overall, panel (d) the share of strong cooperators overall.

the strategies of (cooperating) subject types allow us to resolve these concerns. Figure 3 now plots the cooperation probabilities according to the estimated strategies of cooperating subjects against two predictors of cooperation (expected payoffs and $\delta - \delta^*$). In the left column of plots, we see how the cooperation probabilities across states relate to the difference of discount factor δ and BOS threshold δ^*. In the right column of plots, we see how the probability of cooperation relates to the monetary incentive to cooperate, $\pi_\omega(c) - \pi_\omega(d)$ as defined above, Eqs. 1–3, for each state ω. For these plots, we assume that subjects hold "false consensus" beliefs that their opponent plays the same strategy that they play. That is, strong cooperators believe they face strong cooperators and weak cooperators believe they face weak cooperators. In comparison to Figure 2, the results do not change substantially: Behavior is still close to being independent of the predictors of cooperation in most states (bottom three panels), arguably with the exception of strong cooperators in round 1 (top two panels).

Recall that we know from Dal Bó (2005) and subsequent work that average cooperation rates change as payoff parameters change, and above we have seen that most of these changes can be reduced to changes in round 1. Yet, as just seen, the type strategies are largely inde-

pendent of the payoff parameters. To illuminate this further, we next test the complementary statistic and examine how the shares of the three subject types change as parameters change. Figure 4 plots the shares of cooperators as a function of the discount factor δ in relation to the BOS-threshold δ^*. We see two relatively strong effects: As δ approaches δ^*, the overall share of cooperators increases, i.e. defectors become cooperators, and at the same time, the relative share of cautious cooperators declines, i.e. cautious cooperators turn into strong cooperators.

Result 3 (Question 3 – part 1)**.** *The shares of subjects playing either of the three strategies change highly predictably. As δ increases defectors are replaced by cooperators and as it passes the BOS-threshold δ^* the strong cooperators start to outweigh the cautious cooperators ($\tilde{R}^2 \geq 0.2$ in all cases). The strategies themselves are largely invariant to treatment parameters and monetary incentives. The only exception satisfying $\tilde{R}^2 \geq 0.2$ are strong cooperators in round 1, whose strategies correlate with treatment parameters ($\delta - \delta^*$) but not with monetary incentives, and only in round 1.*

That is, the behavioral changes observed in the literature are mainly transitions from defection to cautious cooperation and from cautious cooperation to strong cooperation. These transitions are neatly predictable, being logistic functions of $\delta - \delta^*$, which is a substantial result in relation to previous work that found no reliable association between strategies used and payoff parameters (Dal Bó and Fréchette, 2018). This result directly follows from the unrestricted estimation of strategies, which thus not only fits better but also renders type shares predictable. In turn, the actual strategies associated with these seemingly archetypical behavioral types are largely invariant of payoff parameters, which also is a novel result that we further investigate next.

Structural analysis of preferences and beliefs The above observations suggest that it seems straightforward to explain the shares of subject types across treatments, as they are simple functions of payoff parameters (i.e. of $\delta - \delta^*$), but it is not immediately obvious how to explain the largely invariant strategies chosen by these subject types. Should we think of cooperating subjects as choosing automata, as first described by Rubinstein (1986), or are these strategies predictable and rationalizable in some way? This question can be answered in a structural analysis of behavior, a standard approach in behavioral analyses of normal-form games, but novel in analyses of the repeated PD. Thanks to the existing work on behavior in normal-form games, we can build on central ideas from three literatures. First, regarding belief formation and relating to the above discussion, much of the pychological literature emphasizes that people overestimate the extent to which others are similar to themselves. In analyses of games, this basic psychological observation has been analyzed as projection of information (Madarász, 2012) and projection of types or strategies (Breitmoser, 2019). By our results above, the types are defined in terms of strategy, and in this sense, projection of types equates with projection of strategies, and both correspond with the false consensus beliefs discussed above. We will contrast these consensus beliefs about opponents' strategies with naive beliefs and Bayesian beliefs in order to test the above suggestion that consensus beliefs best capture behavior (for formal definitions, see Appendix C).

The above results imply, however, that relaxing assumptions on belief formation is insufficient to comprehensively explain behavior. To see this, recall that subjects cooperate after cc and cd/dc even in many treatments where Grim is not a SPE, and if Grim is not a SPE, a strategy involving cooperation is not rationalizable in any state (i.e. never a best response to any belief in any state). Thus, the behavior of cooperating subjects is in general not rationalizable in the classical sense—by varying beliefs—but it may be rationalizable after (also) relaxing assumptions on preferences. A standard approach towards explaining cooperative behavior in the absence of strategic incentives, e.g. if Grim is not a SPE, is to allow for interdependence of preferences. This has been observed in several other literatures, most prominently in finitely repeated public goods games, where such behavior seems related to a preference for conditional cooperation, concerns of inequity aversion, or concerns for fairness and altruism (see for example Keser and Van Winden, 2000, and Fischbacher et al., 2001). Building on this existing evidence, and seeking to avoid post-hoc experimentation, we will only consider these four standard models in our analysis (the standard definitions are provided in Appendix C).

In addition, we allow for the possibility that subjects misperceive the discount factor δ. Such misunderstanding might arise if subjects are used to engaging in repeated interactions with discount factors close to 1 or 0, for example because the most prominent real-life interactions (with say family members and colleagues) have low break-up probability and occur with high frequency, implying that discounting is negligible. Specifically, we allow the perceived discount factor $\tilde{\delta}$ to be a function of the true discount factor as in $\tilde{\delta} = \delta^x$. If $x = 1$, subjects correctly perceive the discount factor (or, break-up probability), for $x < 1$ they underestimate it, with the limiting case $x \to 0$ where they simply disregard the break-up probabibility and play the game as if it had an infinite time horizon (without impatience, in the laboratory). In turn, if $x > 1$, subjects overestimate the break-up probability, and in the limiting case $x \to \infty$, subjects seem "myopic" and play a sequence of one-shot games. Such limitations of foresight characterize many approaches towards long-run interactions, most notably perhaps chess. In the extreme case $\delta \to \infty$, agents simply evaluate the resulting outcome of the present round, i.e. cc, dc, cd or dd, which implicitly encodes the continuation payoff expected from the subsequent rounds.

Regarding the econometric implementation of the analysis, we use standard specifications of structural analyses of games, following McKelvey and Palfrey (1995), Costa-Gomes et al. (2001), Bajari and Hortacsu (2005), as extended to analyses of dynamic games by Aguirregabiria and Mira (2007). All details of these overall standard definitions are provided in Appendix C. In order to quantify to what extent the different approaches allow us to capture behavior, we also estimate two standard benchmark models. First, we provide results for the lower-bound benchmark of *uniform randomization*, i.e. the goodness-of-fit of predicting 50-50 randomization in all states. Second, we consider the upper-bound benchmark *clairvoyance* predicting the actually estimated probabilities of cooperation for the two cooperating types by treatment in all states. Additionally, as a presumably trivial benchmark model, we examine the possibility that subjects play the actual stage game (without interdependent preferences) but misunderstand δ as captured by $x \neq 1$. By design, all models of interdependent preferences allowing for $x \neq 1$ should improve on this benchmark, as it is al-

ways contained as a special case for interdependence weights equal to zero. This benchmark allows us to understand how much can be explained by allowing for misperception of δ on its own. In the estimation x is limited to an upper bound of 100 for viability.

The results are presented in Table 5 and summarized in Figure 1 above. Table 5 distinguishes, for each model, three sets of estimates. This gives us a sense of the robustness of the results. In the right-most columns ("Fit to each treatment"), we allow for treatment-specific parameters. Since the behavior of cooperating subjects in each treatment is described by five parameters (round-1 behavior of each type and three parameters capturing continuation behavior), the four free parameters per model, when allowed to be treatment specific, should capture behavior close to "clairvoyance" (i.e., perfectly). This is indeed the case for models allowing for false-consensus beliefs but not for the other belief models, as discussed shortly.

In the the middle set of columns ("Heterogeneous variance"), we allow for treatment-specific variance of noise but now invoke the standard assumption that the preference parameters are constant across all treatments and experiments, while the clairvoyance benchmark model remains unchanged (aside from a change in the penalty term to reflect the change in the number of free parameters of the models for which it is the upper bound). This informs us to what extent interdependence of preferences actually captures behavior, rather than being able to fit behavior post-hoc treatment by treatment. In the left-most set of columns ("Homogeneous variance"), we additionally assume that the noise variance (as captured by the precision parameter λ in the logistic specification) is constant across treatments. This yields a very parsimonious model of behavior, using four parameters to describe the strategies used of both cooperative types across all 32 treatments analyzed here—which may not be expected to fit exactly. Explaining behavior across treatments and experiments with one set of parameters gives us a sense of how robust (and thus predictable) behavior is, however.

As indicated, for each belief model, we evaluate the aforementioned four models of social preferences with potentially misperceived δ, and a benchmark of inequity aversion assuming the correct δ. Two observations stand out: First, for each of the models with interdependent preferences, and each of the three measures for the goodness-of-fit, false consensus beliefs best fit behavior—that is, cautious cooperators seem to believe they play against cautious cooperators and strong ones seem to believe they play against strong ones. In all cases, the distance to other belief models is on the order of 5000 likelihood points, which is highly significant and corresponds to about 20% of the total score, implying that it is behaviorally also highly relevant.

To understand this first observation, let us assume that subjects update beliefs following Bayes' Rule after each round, most notably perhaps after round 1—which could explain the poor fit of actions in relations to expected payoffs after round 1. Since all *cooperating* subjects are estimated to play the same continuation strategy, their differences in round 1 cannot be in preferences but must be in the beliefs they hold, and specifically in the beliefs about behavior in round 1, as the behavioral differences are observed in round 1. False consensus about strategies directly predicts this intuition— that cautious cooperators expect to play with cautious cooperators and that strong ones expect to play with strong ones—and it also reflects the standard theoretical assumption that agents play symmetric equilibrium strategies.

Table 5: Testing interdependence of preferences (second halves; see also Tables 10 and 11 for analyses of first halves and both halves)

	Fit to pooled data					
	Homogeneous variance		Heterogeneous variance		Fit to each treatment	
Model (free parameters)	*BIC*	Estimates	*BIC*	Estimates	*BIC*	Average Estimates
Upper bound BIC (Clairvoyance)	20460.6		20692.6		21388.8	
Lower bound BIC (Uniform Random)	51487.3		51719.4		52415.6	
False Consensus Beliefs						
True supergame (g,l,δ), no free par $(-)$	45115.4	$(-,-,-)$	42523.4	$(-,-,-)$	43219.6	$(-,-,-)$
True stage game g,l, free $(\delta^X,-,-)$	45096.6	$(1.08,-,-)$	42134	$(1.32,-,-)$	39948.6	$(8.79,-,-)$
True δ, inequity aversion $(-,\alpha,\beta)$	28542.4	$(-,0.96,0.6)$	29407.7	$(-,1.6,0.66)$	27950.4	$(-,-100,0.52)$
Inequity Aversion (δ^X,α,β)	22607.6	$(100,0.82,0.14)$	22330.2	$(18.44,0.77,0.11)$	21452.4	$(17.05,0.37,-0.01)$
Cond Cooperation (δ^X,α,β)	27159.5	$(100,1.61,-0.27)$	25680.3	$(5.91,1.7,-0.01)$	21767.4	$(16.79,1.79,-0.06)$
Altruism (δ^X,α,β)	24309.4	$(68.15,1.45,-0.32)$	23419.6	$(19.92,1.38,-0.24)$	21451.1	$(4.17,0.98,0.12)$
Gen Fairness Equilibrium (δ^X,α,β)	28525.3	$(6.53,6.66,0.22)$	26864.2	$(6.75,26.51,0.21)$	22067.6	$(11.03,24.23,0.07)$
Naive Beliefs						
True supergame (g,l,δ), no free par $(-)$	44692.6	$(-,-,-)$	43458.3	$(-,-,-)$	44154.5	$(-,-,-)$
True stage game g,l, free $(\delta^X,-,-)$	44638.9	$(1.08,-,-)$	43310.6	$(1.14,-,-)$	41986.3	$(2.53,-,-)$
True δ, inequity aversion $(-,\alpha,\beta)$	31003	$(-,-100,-3.27)$	31175.5	$(-,-100,-2.18)$	30032.2	$(-,-100,-2.34)$
Inequity Aversion (δ^X,α,β)	27869.8	$(100,6.99,0.98)$	27782.3	$(100,4.57,0.98)$	28007.6	$(100,8.48,0.81)$
Cond Cooperation (δ^X,α,β)	34743.7	$(100,5.88,0.03)$	31846.3	$(4.73,3.3,0.28)$	28008.3	$(99.06,4.76,-0.12)$
Altruism (δ^X,α,β)	29473.8	$(100,33.58,-0.8)$	28683.2	$(20.19,4.48,-0.7)$	28008.3	$(100,12.53,-0.54)$
Gen Fairness Equilibrium (δ^X,α,β)	29630.5	$(4.64,-8.1,0.53)$	28729.9	$(4.11,-5.92,0.53)$	28008.3	$(34.26,-10.75,0.54)$
Bayesian Beliefs						
True supergame (g,l,δ), no free par $(-)$	44424.9	$(-,-,-)$	42421.6	$(-,-,-)$	43117.8	$(-,-,-)$
True stage game g,l, free $(\delta^X,-,-)$	44022	$(0.78,-,-)$	42342	$(0.89,-,-)$	41036.3	$(10.09,-,-)$
True δ, inequity aversion $(-,\alpha,\beta)$	31871.5	$(-,2.15,0.93)$	33302.6	$(-,2,0.65)$	33004.8	$(-,100,100)$
Inequity Aversion (δ^X,α,β)	28095.3	$(100,5.71,0.81)$	28091.1	$(74.82,16.36,0.79)$	28508.4	$(30.89,15.78,0.87)$
Cond Cooperation (δ^X,α,β)	35378.4	$(100,3.53,-0.11)$	32160.8	$(3.85,2.04,0.12)$	28501.6	$(1,5.66,-0.3)$
Altruism (δ^X,α,β)	29162	$(100,-50.45,5.08)$	28915.4	$(14.07,-17.8,4.89)$	28505.3	$(34.72,54.88,-0.3)$
Gen Fairness Equilibrium (δ^X,α,β)	34577.4	$(5.58,11.59,0.17)$	32527	$(5.69,11.59,0.15)$	28505.3	$(23.52,100,-0.11)$

Note: This table shows the estimates and BICs for the estimated models including benchmarks. In the rightmost column ("Fit to each treatment") parameters are estimated by treatment – the BICs are aggregated, the reported parameter estimates are averages. In the columns ("Fit to pooled data") parameter sets are estimated to be constant across all experiments with homogeneous variance and heterogeneous variance by treatment, respectively. The upper bound and lower bound BIC are based on the same "Clairvoyance" and "Uniform Random" model in all three columns, with treatment specific strategies for the clairvoyance model, but the BICs take into account the differences in parameter numbers of the interdependent-preferences models across the three columns to make them comparable.

Second, between the four well-known models of interdependent preferences (detailed in the Appendix C), there is a clear ranking when applied to the range of experiments we re-analyze here. Whatever assumption we impose on the belief model, capturing interdependence by inequity aversion fits substantially and significantly better than capturing interdependence by any other model. That is, we observe very robust rankings of models with respect to both dimensions, beliefs and preferences. We attribute this to the comprehensive data set re-analyzed here, which reduces the impact of single observations and allows the law of large numbers to take effect.

Result 4 (Question 3 – part 2)**.** *Subjects' behavior is best described by false consensus beliefs (i.e. symmetric equilibrium) and inequity aversion. Indeed, false consensus fits substantially better than other belief models for all models of interdependent preferences, and inequity aversion equally fits substantially better than other interdependence models for all belief models.*

Next, let us look at the extent of misperception of δ. In total, we consider four models of interdependent preferences, three models of belief formation, and three specifications of treatment dependence of parameters. Between these $36 = 4 \times 3 \times 3$ sets of estimates, we obtain 35 times an estimate indicating $x > 1$, i.e. $\delta^x < \delta$, and in particular, this is true for the identified specifications where subjects either hold false consensus beliefs or exhibit inequity aversion. Indeed, when we allow subjects to both hold false consensus beliefs and exhibit inequity aversion, and in many other cases, we estimate the upper bound $x = 100$, implying $\delta^x \approx 0$. Thus, subjects are clearly best described by limited foresight, similar to (but much more extreme than) the chess players referenced above: Given $\delta^x \approx 0$, subjects in the repeated PD do not seem to look beyond the current round. They capture the expected payoffs from continuation play by the values they associate with each of the four possible outcomes (cc, dc, cd, dd) of play in the current round, and these values relate to the stage game payoffs via inequity aversion.

Result 5 (Question 3 – part 3)**.** *Subjects are estimated to not look ahead beyond the present round, and the state values they associate with the outcome of play in the present round relate to the stage game payoffs via inequity aversion.*

So, which types of games are induced by the state values perceived by the subjects? The answer depends on the stage game payoffs in the respective treatments, but to give some sense, let us look at two well-known examples.

	c	d
c	$2,2$	$0,3$
d	$3,0$	$1,1$

$\overset{\rightsquigarrow}{\alpha=0.82,\beta=0.14}$

	c	d
c	$2,2$	$-0.42,0.54$
d	$0.54,-0.42$	$1,1$

	c	d
c	$3,3$	$0,4$
d	$4,0$	$1,1$

$\overset{\rightsquigarrow}{\alpha=0.82,\beta=0.14}$

	c	d
c	$3,3$	$-0.56,0.72$
d	$0.72,-0.56$	$1,1$

It is easy to verify that for a wide range of stage game payoffs, inequity aversion with the estimated parameters $(0.82, 0.14)$ induces a coordination game. Formally, a coordination game is obtained if $g < \alpha * (1 + g + l)$, and using $\alpha = 0.82$, this holds true whenever $g \leq 4$, which includes all of the experimental games we analyze. That is, in terms of the continuation payoffs, subjects generally seem to perceive the repeated PD as a coordination game. Being a coordination game, there exist three Nash equilibria – the defective equilibrium (d,d), the cooperative equilibrium (c,c), and a mixed equilibrium corresponding to $\Pr(c) = 0.49$ in the upper game and to $\Pr(c) = 0.41$ in the lower game. So, how does the econometric model align subjects' behavior with play this coordination game? After round 1, subjects play the "Schelling points" of the coordination game (Schelling, 1960), i.e. the focal point given by the previous round's choices, and they correspondingly play the cooperative, defective or mixed equilibrium after cc, dd, and cd/dc, respectively. In round 1, there is no such focal point, and subjects focus on either the cooperative, or the mixed, or the defective equilibrium, depending on subjective beliefs and yielding the three subject types observed above (strong cooperators, cautious cooperators, and defectors, respectively). As demonstrated, the type shares (i.e. the subjective beliefs) depend in a clear-cut way on the game parameters, and as we also saw by the significance of the type distinction, at the subject level the focus is robust. To clarify, if it were not robust at the subject level, then the distinction of say cautiously and strongly cooperative subjects would not have been found to be significant.

6 Conclusion

We summarize our main results as follows.

Re-analyzing 12 experiments, we robustly identify three different types of subjects: defectors, playing a strategy close to AD, and cautious and strong cooperators who play semi-grim strategies that differ in their first-round cooperation probability. The strategies are largely independent of treatment parameters but the shares of subjects picking either of the three strategies depend strikingly on the continuation probability δ in relation to the BOS-threshold δ^* (Blonski et al., 2011). Following rounds where at least one player cooperated, subjects cooperate systematically even in supergames where Grim is not a subgame-perfect equilibrium, which is rationalizable after allowing for interdependent preferences. Testing different belief and interdependent preference models in a structural analysis, we find that the observed behavior can be explained by subjects holding false-consensus beliefs, and having limited foresight as well as inequity-averse preferences.

Specifically, subjects are estimated to play each round of the repeated PD based on subjective valuations of the states that will result from the current round's choices. These state values relate to the original stage game payoffs in a manner compatible with inequity aversion and induce coordination games for the experimental games we consider. The defectors play according to the defective equilibrium in round 1 and thereafter. Some of the cooperating subjects systematically play according to the cooperative equilibrium in round 1 and are identified as strong cooperators, while the others systematically playing according to the

mixed equilibrium and are identified as cautious cooperators. This focus in round 1 is persistent at the subject level. In the subsequent rounds, both types of cooperative subjects play the Schelling points, i.e. according to the cooperative equilibrium after (c,c), according to the defective equilibrium after (d,d), and according to the mixed equilibrium after $(c,d)/(d,c)$.

This description of behavior in the repeated PD is the result of a flexible structural analysis of 12 experiments, it closely relates to a wide range of previous results in behavioral economics, and it fits behavior very well also quantitatively (see Figure 1). Using merely four parameters to explain 65.910 and 79.892 observations of inexperienced and experienced subjects (respectively), it captures 89% of the variance in behavior of inexperienced subjects and 93% of behavior of experienced subjects from 32 treatments. The results also connect with key results in several large strands of the literature. False consensus is a central concept in psychology (Ross et al., 1977), the idea that the actions in the previous round serve as focal point for the actions in the present round is (informally) predicted by the focal point theory (Schelling, 1960), limited foresight and state recognition/evaluation are central ideas in games with indefinite time-horizon in computer science (Levy and Newborn, 1982), in economics (Jehiel, 2001; Kübler and Weizsäcker, 2004), and even for grand-master chess players (Gobet and Simon, 1996), and inequity aversion (Fehr and Schmidt, 1999) is a central concept of interdependent preferences. Further, we can rule out many potential confounds related to overfitting when a model with four parameters explains 93% of variance from close to 80.000 observations that were taken in a wide range of conditions.

The observations that subjects assign values to future states and that the state values closely relate to stage game payoffs in a manner compatible with inequity aversion are very encouraging for future work, and perhaps most importantly, they represent a first behavioral foundation of play in repeated games—i.e. a formally closed explanation of behavior that enables predictions for all repeated games. Experimental work on repeated games other than the repeated PD is needed to evaluate these predictions, but the observation that closed behavioral models, and structural analyses such as those known from static games, are possible also for repeated games demonstrate that it is feasible and important to move beyond strategy estimation in attempts towards understanding behavior. In addition, our results raise a number of novel and interesting questions with respect to analyses of the repeated PD. We considered inequity aversion mainly because it is a well-established model of interdependent preferences used to explain cooperative behavior in prior work. Thinking of state values, should we not also include the true discount factor δ as a relevant factor? Is it a coincidence that the recurring semi-grim strategies are specific instances of belief-free equilibria (Ely et al., 2005)? Is their apparent invariance after round 1 not reminiscent also of analogical reasoning (Samuelson, 2001)? Over time, behavior in round 1 seems to somewhat change as subjects gain experience (Fudenberg and Karreskog, 2020)—though the changes cancel out across treatments (see Table 2)—does the "precision" λ change, do beliefs change, or do preferences change? Following the approach towards structurally analyzing behavior in repeated games developed above, it will be possible to ask and answer these and many more such questions in exciting future work.

References

Aguirregabiria, V. and Mira, P. (2007). Sequential estimation of dynamic discrete games. *Econometrica*, 75(1):1–53.

Ansari, A., Montoya, R., and Netzer, O. (2012). Dynamic learning in behavioral games: A hidden markov mixture of experts approach. *Quantitative Marketing and Economics*, 10(4):475–503.

Aoyagi, M. and Frechette, G. (2009). Collusion as public monitoring becomes noisy: Experimental evidence. *Journal of Economic theory*, 144(3):1135–1165.

Bajari, P. and Hortacsu, A. (2005). Are structural estimates of auction models reasonable? evidence from experimental data. *Journal of Political Economy*, 113(4):703–741.

Biernacki, C., Celeux, G., and Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE transactions on pattern analysis and machine intelligence*, 22(7):719–725.

Bilmes, J. A. et al. (1998). A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. *International Computer Science Institute*, 4(510):126.

Blonski, M., Ockenfels, P., and Spagnolo, G. (2011). Equilibrium selection in the repeated prisoner's dilemma: Axiomatic approach and experimental evidence. *American Economic Journal: Microeconomics*, 3(3):164–192.

Breitmoser, Y. (2015). Cooperation, but no reciprocity: Individual strategies in the repeated prisoner's dilemma. *American Economic Review*, 105(9):2882–2910.

Breitmoser, Y. (2019). Knowing me, imagining you: Projection and overbidding in auctions. *Games and Economic Behavior*, 113:423–447.

Breitmoser, Y., Tan, J. H., and Zizzo, D. J. (2014). On the beliefs off the path: Equilibrium refinement due to quantal response and level-k. *Games and Economic Behavior*, 86:102–125.

Brunner, C., Camerer, C. F., and Goeree, J. K. (2011). Stationary concepts for experimental 2 x 2 games: Comment. *American Economic Review*, 101(2):1029–40.

Bruttel, L. and Kamecke, U. (2012). Infinity in the lab. how do people play repeated games? *Theory and Decision*, 72(2):205–219.

Camera, G., Casari, M., and Bigoni, M. (2012). Cooperative strategies in anonymous economies: An experiment. *Games and Economic Behavior*, 75(2):570–586.

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

Costa-Gomes, M., Crawford, V. P., and Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 69(5):1193–1235.

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

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

Dal Bó, P. and Fréchette, G. R. (2015). Strategy choice in the infinitely repeated prisoners' dilemma. *Working paper.*, Available at SSRN: https://ssrn.com/abstract=2292390 or http://dx.doi.org/10.2139/ssrn.2292390.

Dal Bó, P. and Fréchette, G. R. (2018). On the determinants of cooperation in infinitely repeated games: A survey. *Journal of Economic Literature*, 56(1):60–114.

Dreber, A., Fudenberg, D., and Rand, D. G. (2014). Who cooperates in repeated games: The role of altruism, inequity aversion, and demographics. *Journal of Economic Behavior & Organization*, 98:41–55.

Dreber, A., Rand, D. G., Fudenberg, D., and Nowak, M. A. (2008). Winners don't punish. *Nature*, 452(7185):348–351.

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

Dvorak, F. and Fehrler, S. (2018). Negotiating cooperation under uncertainty: Communication in noisy, indefinitely repeated interactions. *Working paper*.

El-Gamal, M. A. and Grether, D. M. (1995). Are people bayesian? uncovering behavioral strategies. *Journal of the American statistical Association*, 90(432):1137–1145.

Ely, J. C., Hörner, J., and Olszewski, W. (2005). Belief-free equilibria in repeated games. *Econometrica*, 73(2):377–415.

Eyster, E. and Rabin, M. (2005). Cursed equilibrium. *Econometrica*, 73(5):1623–1672.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics letters*, 71(3):397–404.

Fréchette, G. R. and Yuksel, S. (2017). Infinitely repeated games in the laboratory: Four perspectives on discounting and random termination. *Experimental Economics*, 20(2):279–308.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Science & Business Media.

Fudenberg, D. and Karreskog, G. (2020). Predicting average cooperation in the repeated prisoner's dilemma. *Working Paper*.

Fudenberg, D. and Maskin, E. S. (1986). Folk theorem for repeated games with discounting or with incomplete information. *Econometrica*, 54(3):533–554.

Fudenberg, D., Rand, D. G., and Dreber, A. (2012). Slow to anger and fast to forgive: Cooperation in an uncertain world. *American Economic Review*, 102(2):720–749.

Gobet, F. and Simon, H. A. (1996). The roles of recognition processes and look-ahead search in time-constrained expert problem solving: Evidence from grand-master-level chess. *Psychological science*, 7(1):52–55.

Goeree, J. K., Holt, C. A., and Palfrey, T. R. (2003). Risk averse behavior in generalized matching pennies games. *Games and Economic Behavior*, 45(1):97–113.

Harless, D. W. and Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, 62(6):1251–1289.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70.

Houser, D., Keane, M., and McCabe, K. (2004). Behavior in a dynamic decision problem: An analysis of experimental evidence using a bayesian type classification algorithm. *Econometrica*, 72(3):781–822.

Houser, D. and Winter, J. (2004). How do behavioral assumptions affect structural inference? evidence from a laboratory experiment. *Journal of Business & Economic Statistics*, 22(1):64–79.

Imhof, L. A., Fudenberg, D., and Nowak, M. A. (2007). Tit-for-tat or win-stay, lose-shift? *Journal of theoretical biology*, 247(3):574–580.

Jehiel, P. (2001). Limited foresight may force cooperation. *The Review of Economic Studies*, 68(2):369–391.

Kagel, J. H. and Schley, D. R. (2013). How economic rewards affect cooperation reconsidered. *Economics Letters*, 121(1):124–127.

Keser, C. and Van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *scandinavian Journal of Economics*, 102(1):23–39.

Kübler, D. and Weizsäcker, G. (2004). Limited depth of reasoning and failure of cascade formation in the laboratory. *The Review of Economic Studies*, 71(2):425–441.

Levy, D. and Newborn, M. (1982). How computers play chess. In *All About Chess and Computers*, pages 24–39. Springer.

Madarász, K. (2012). Information projection: Model and applications. *The Review of Economic Studies*, 79(3):961–985.

McKelvey, R. D. and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38.

Nowak, M. and Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, 364(6432):56.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American economic review*, pages 1281–1302.

Rand, D. G., Fudenberg, D., and Dreber, A. (2015). It's the thought that counts: The role of intentions in noisy repeated games. *Journal of Economic Behavior & Organization*, 116:481–499.

Romero, J. and Rosokha, Y. (2019). Mixed strategies in the indefinitely repeated prisoner's dilemma. *Available at SSRN 3290732*.

Ross, L., Greene, D., and House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of experimental social psychology*, 13(3):279–301.

Rubinstein, A. (1986). Finite automata play the repeated prisoner's dilemma. *Journal of economic theory*, 39(1):83–96.

Samuelson, L. (2001). Analogies, adaptation, and anomalies. *Journal of Economic Theory*, 97(2):320–366.

Schelling, T. C. (1960). *The Strategy of Conflict*. Harvard university press.

Schennach, S. M. and Wilhelm, D. (2017). A simple parametric model selection test. *Journal of the American Statistical Association*, pages 1–12.

Shachat, J., Swarthout, J. T., and Wei, L. (2015). A hidden markov model for the detection of pure and mixed strategy play in games. *Econometric Theory*, 31(4):729–752.

Sherstyuk, K., Tarui, N., and Saijo, T. (2013). Payment schemes in infinite-horizon experimental games. *Experimental Economics*, 16(1):125–153.

Stahl, D. O. and Wilson, P. W. (1994). Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization*, 25(3):309–327.

Weizsäcker, G. (2003). Ignoring the rationality of others: evidence from experimental normal-form games. *Games and Economic Behavior*, 44(1):145–171.

Wright, S. P. (1992). Adjusted p-values for simultaneous inference. *Biometrics*, 48:1005–1013.

Online appendix

Inequity Aversion and Limited Foresight in the Repeated Prisoner's Dilemma

A Econometric approach toward strategy estimation

Recall that a subject using a pure strategy acts equivalently whenever a given state is reached and she uses the same pure strategy across all supergames. A subject using a mixed strategy uses a pure strategy within supergames but randomizes over pure strategies prior to supergames. A subject using a behavior strategy may randomize each round and thus deviate from pure strategies even within supergames. These definitions provide a basis for identification, but identification is made difficult by the standard assumption that choice is stochastic. For example, a single deviation from a given pure strategy, over say 20 observations, is intuitively not considered sufficient evidence against purity of strategies. Otherwise, the case for behavior strategies would be trivial, but how can this intuition be made formally precise—in a manner that allows us to econometrically distinguish "noisy" pure, mixed, and behavior strategies?

The distinction is achieved efficiently using the Markov-switching models known from empirical finance and empirical macroeconomics in conjunction with the robust likelihood-ratio tests of Schennach and Wilhelm (2017). Markov-switching models generalize the finite-mixture and random-switching models used in previous analyses of repeated game strategies.[14] They allow us to capture a potentially heterogeneous group of agents (in our case, subjects potentially playing different strategies), where each agent is characterized by a "state of mind" (the strategy to be played), and agents may change their states of mind over the course of time, but both states and transitions are latent and thus not directly observable. Let us refer to Ansari et al. (2012), Breitmoser et al. (2014) and Shachat et al. (2015) for earlier applications in behavioral analyses. The identifying assumption is that state transitions follow a Markov process. This generalizes the finite mixture model, with degenerate transition probabilities, and the random switching model, with constant choice probabilities for the strategies. Given this, estimation proceeds by maximum likelihood using an EM algorithm. Model adequacy is evaluated using ICL-BIC (Biernacki et al., 2000), and model differences are evaluated using the Schennach-Wilhelm test, which captures that all models may be arbitrarily nested and misspecified. Finally, we allow for stochastic choice in the form of trembles (after all histories of play) following Harless and Camerer (1994), i.e. in

[14] The approach of using mixture models in order to uncover decision rules in experimental data has been established by Stahl and Wilson (1994) and El-Gamal and Grether (1995) and subsequently used in many analyses of level-k reasoning and stochastic choice, see e.g. Houser and Winter (2004) and Houser et al. (2004), to unravel individual decision rules. A special case of finite mixture modeling is the Strategy Frequency Estimation Method (SFEM) employed by Dal Bó and Fréchette (2011), Fudenberg et al. (2012), Rand et al. (2015), Dal Bó and Fréchette (2015), Fréchette and Yuksel (2017).

each round the minimal probability of any action is equal to $\gamma \geq 0$ where γ is a free (noise) parameter in the estimation.

A.1 Markov-Switching Models and ICL-BIC

The Markov-switching model builds on the simpler and more restrictive finite mixture model, which has been established in the experimental literature by Stahl and Wilson (1994) and can be used to empirically identify a finite number K of strategies with parameter vectors θ_k. The log-likelihood function to be maximized for the finite mixture model is

$$\ln \mathcal{L}(\theta, \rho|O) = \log \left(\prod_{s \in S} p(o_s|\theta, \rho) \right) = \sum_{s \in S} \ln \sum_{k \in K} \rho_k \, p_k(o_s|\theta_k), \tag{4}$$

with observations O, ρ_k denoting the relative frequency of strategy k, and $p_k(o_s|\theta_k)$ denoting the probability that player s chooses action o_s given he plays strategy k [15].

A way to model regime switches is to replace the implicit latent indicator variable in finite mixture models (indicating the discrete types) with a hidden Markov chain (Frühwirth-Schnatter, 2006). The central assumption characterizing the learning process in Markov models is that the type of a player (or its strategy in our context) in the next period can only depend on its type in this period. More precisely if k_t is the type in period t then: $Pr(k_{t+1}|k_t, k_{t-1}, k_{t-2}, ..., k_1) = Pr(k_{t+1}|k_t)$, where the type is hidden and cannot be observed directly.[16] What we do observe is the action o_t, which in turn depends on the type k_t in t only: $Pr(o_t|k_t, o_{t-1}, k_{t-1}, ..., k_1, o_1) = Pr(o_t|k_t)$ (c.f Bilmes et al. (1998)). It implies that transitions between states are independent of time t. This assumption might be quite restrictive. For example if we want to assume that the probability of switching to a new strategy is more likely later in the game than at the beginning. Nevertheless, we can use memory-2 or memory-3 strategies if we define the state ω as a history of more than one past outcome and condition the strategy on this history of outcomes. Moreover, it is possible to interact time dependent components with switching probabilities.

Let K_t denote the state at time $t \in 1, 2, ..., T$ and $\sigma_{kk'} = Pr(K_{t+1} = k'|K_t = k)$ define the transition probability from k to k' which is independent from t, as pointed. So σ is a $(K \times K)$ transition matrix containing transition probabilities for every pair of states, where all entries are positive and each row sums up to 1. Moreover, the state-paths are denoted by $\kappa \in K^T$ with $Pr(\kappa)$ conditional on initial weights ρ and transition probabilities σ. The probability of observing $o_{s,t}$ conditional on subject s being type k in this period is $Pr(o_{s,t}|\theta, k)$. The likelihood function is:

$$\ln \mathcal{L}(\rho, \sigma, \theta|O) = \sum_{s \in S} \ln \sum_{\kappa \in K^T} Pr(\kappa) \prod_{t \leq T} Pr(o_{s,t}|\theta, \kappa(t), t) \tag{5}$$

[15] For the memory-1 case $p_k(o_s|\theta_k) = \prod_t (\sigma_{\omega_{s,t}}(k))^{o_{s,t}} (1 - \sigma_{\omega_{s,t}}(k))^{1-o_{s,t}}$ with strategy $\sigma_{\omega_{s,t}}(k)^{o_{s,t}}$ for state $\omega_{s,t}(k)^{1-o_{s,t}}$.

[16] Therefore also known as the Hidden Markov Model (HMM).

Due to the introduction of the transition matrix σ the number of parameters to be estimated increases dramatically. Moreover, with a naive estimation approach we would have to consider all possible state paths and be very time consuming. Therefore we choose to apply a backward-forward algorithm to calculate posteriors for estimation with the expectation maximization (EM) algorithm.

The idea of the EM-algorithm is to conduct two steps the E-step and the M-step iteratively. This way we split up every optimization step into many steps which simplifies complexity and consequently speeds up computations. In the E-step we evaluate the conditional expectation of the log-likelihood given our data O and the current parameter vector and then maximize over a reduced set of free parameters in the M-Step. The number of possible types k is pre-defined as well as the structure of their mixing parameters θ_k.

In the E-step we need to compute for all subjects for all time periods the posterior probability of component inclusion (being a specific type) and the probability to switch between two types. An efficient way to calculate those posterior probabilities is to built up on the backward-forward. First, we have the forward procedure, where we define the (joint) probability of observing the partial sequence $o_{s1},...,o_{st}$ and ending up with type k at time t:

$$\alpha_{sk}(t) = Pr(O_{s1} = o_{s1},...,O_{st} = o_{st}, K_t = k) \tag{6}$$

Recursively, we can then define:

$$\begin{aligned} &1. \quad \alpha_{sk}(1) = \rho_k Pr(o_{s1}|\theta,k) \\ &2. \quad \alpha_{sk'}(t+1) = \left[\sum_k \alpha_{sk}(t)\sigma_{kk'}\right] Pr(o_{st+1}|\theta,k') \\ &3. \quad Pr(o_s) = \sum_{k\in K} \alpha_{sk}(T) \end{aligned} \tag{7}$$

Second, for the backward procedure we define the probability of ending in the partial sequence $o_{st+1},...o_{sT}$ given that we have started at type k at time t.

$$\beta_{sk}(t) = Pr(O_{st+1} = o_{st+1},...,O_T = o_T|K_t = k) \tag{8}$$

Again we can define $\beta_{sk}(t)$ efficiently (Bilmes et al., 1998)

$$\begin{aligned} &1. \quad \beta_{sk}(T) = 1 \\ &2. \quad \beta_{sk}(t) = \sum_{k'\in K} \sigma_{kk'} Pr(o_{st+1}|\theta,k')\beta_{sk'}(t+1) \\ &3. \quad Pr(o_s) = \sum_{k\in K} \beta_{sk}(1)\rho_k Pr(o_{s1}|\theta,k) \end{aligned} \tag{9}$$

We then take advantage of the fact that the unconditional probability $Pr(o_s)$ can be defined using $\alpha_{sk}(t)$ or $\beta_{sk}(t)$ to calculate the posterior probabilities γ_{sk} and $\zeta_{skk'}$. The former

is the conditional probability of being type k at time t given observations o_s:

$$\gamma_{sk}(t) = Pr(K_t = k|o_s) = \frac{Pr(o_s, K_t = k)}{Pr(o_s)} = \frac{Pr(o_s, K_t = k)}{\sum_{k' \in K} Pr(o_s, K_t = k)} = \frac{\alpha_{sk}(t)\beta_{sk}(t)}{\sum_{k' \in K} \alpha_{sk'}(t)\beta_{sk'}(t)}, \tag{10}$$

Using γ_{sk} we can define the probability of having type k in t and type k' in $t+1$ conditional on our observations as

$$\zeta_{skk'}(t) = Pr(K_t = k, K_{t+1} = k'|o_s) = \frac{Pr(K_t = k|o_s)Pr(o_{t+1,\ldots,T}, K_{t+1} = k'|K_t = k')}{Pr(o_{t+1,\ldots,T}|K_t = k)} \tag{11}$$

$$= \frac{\gamma_{sk}(t)\sigma_{kk'}Pr(o_{s,t+1}|\theta, k')\beta_{sk'}(t+1)}{\beta_{sk}(t)}$$

(cf. Bilmes et al. (1998)).

In the M-step we maximize for each k and $t \leq T$ the function

$$LL_{kt}(\theta'_k) = \sum_{s \in S} \gamma_{sk}(t) \ln Pr(o_{st}|\theta') \rightarrow \max_{\theta'_{kt}}! \tag{12}$$

to yield the updated θ^{+1} when assuming that θ_{kt} does not affect the likelihood of other components k. If it does, we need to maximize $\sum_{k' \in K} LL_{kt}(\theta') \rightarrow \max_{\theta'}!$ and yield θ^{+1}.[17] Moreover, we update ρ and σ using the posteriors from above and yield

$$\rho_k^{+1} = \frac{1}{n}\sum_{s \in S} \gamma_{sk}(1) \quad \text{and} \quad \sigma_{kk'}^{+1} = \frac{\sum_{s \in S}\sum_{t<T} \zeta_{skk'}(t)}{\sum_{s \in S}\sum_{t<T} \gamma_{sk}(t)} \tag{13}$$

The two steps are iterated until the distance between (θ, ρ, σ) and $(\theta^{+1}, \rho^{+1}, \sigma^{+1})$ gets small.

Estimation proceeds by a maximum likelihood, as usual, but as is well-known, the larger the number of parameters, the larger a model's capacity to fit data—and implicitly, the larger its fallacy to overfit the data. This is conventionally captured by evaluating model adequacy based on information criteria such as BIC, which penalize for the degrees of freedom in a theoretically adequate manner. Mixture and switching models additionally contain freedom in defining the components of the subject pool, i.e. the number of subject types, which provides an additional source for overfitting aside from the number of parameters used. Following (Biernacki et al., 2000), these concerns are addressed using the information-classification likelihood Bayes-information criterion (ICL-BIC), a criterion that penalizes both model complexity and the failure of the mixture model to provide a classification in well-separated strategy clusters. We address the observation that modeling mixtures of pure, mixed, and behavior strategies induces sophisticated nesting structures, and the concern that indeed all models may be misspecified by evaluating model differences using the novel Schennach-Wilhelm likelihood ratio tests (Schennach and Wilhelm, 2017). Finally, we capture the intuition that

[17] θ may depend on t but does not have to.

choice is stochastic by allowing for trembles in the sense of Selten (1975): Each agent of a player picks any given action with probability no less than e > 0. This approach follows (Breitmoser, 2015) and, in relation to the logistic-error approach proposed by (Dal Bó and Fréchette, 2011), it has the advantage that it does not perturb choice probabilities of subjects that originally randomize already.

A.2 Validity

To demonstrate the validity of our approach to distinguish pure, mixed, and behavior strategies, we first run it on different sets of simulated data: For each of the three conjectures, we simulate corresponding data sets and verify if we can identify the underlying conjecture based on model-fit evaluations using ICL-BIC. As for pure strategies, we consider a population where AD, Grim, and TFT have share 0.25 each, AC has share 0.15, and WSLS has share 0.1.[18] Drawing from this population, we simulate for three different discount factors $\delta = 0.6$, $\delta = 0.75$, and $\delta = 0.9$ each 100 data-sets with 50 subjects[19] and enough supergames to have 40 decisions per subject past round 1.

Here, $\delta = 0.75$ corresponds to the average supergame in our sample, $\delta = 0.6$ and $\delta = 0.9$ serve as robustness check approaching the upper and lower bound of δ in our data. The tremble parameter is $\gamma = 0.1$, which is of the proportion typically estimated in the literature. Then we determine the average ICL-BICs of the three basic econometric models, finite-mixture, random-switching, and semi-grim[20], across those 100 data-sets and compare their performance using simple matched-pairs Wilcoxon tests of the ICL-BICs. Table 6 reports the results.

Under the pure-strategy conjecture, the true model of the population is the finite-mixture model. Our analysis should therefore identify it as the best fitting model if and only if the simulated subjects play pure strategies. The first three rows of Table 6 show, that this is clearly the case: We obtain significantly (at $\alpha = 0.01$) lower ICL-BICs for the finite-mixture model than for the other two models for all three values of δ. We can therefore identify pure strategies with our approach.

We repeat the same exercise for simulated subject pools playing mixed strategies and pools playing semi-grim strategies. The mixed strategy population is based on the same pure strategies and prior probabilities as above but assuming subjects redraw a pure strategy prior to each supergame. The semi-grim strategy is of the form $(0.4, 0.9, 0.3, 0.3, 0.1)$, which approximates the average cooperation probabilities across all experiments in our data set.

The results displayed in the bottom rows of Table 6 indicate that distinguishing between mixed-strategy and semi-grim populations is more difficult. When analyzing long supergames ($\delta = 0.9$), there appears be to a bias towards detecting semi-grim, and analyzing

[18] We include WSLS here and in the analysis below, as a number of studies established its evolutionary robustness, see Nowak and Sigmund (1993) and Imhof et al. (2007), indicating that it should be considered a promising candidate.

[19] Robustness-checks with 100 and 200 subjects are provided in Table 7.

[20] In this case, without allowing for subject heterogeneity, the semi-grim model simply determines the average cooperation rates in each state.

Table 6: Can we econometrically distinguish pure, mixed and behavior strategies?

	Model fitted to the data				
	Finite Mixture		Random Switching		Semi-Grim
Pure-Strategy Conjecture					
$\delta = 0.6$	**1227.92**	$\ll$	1706.67	$\ll$	1862.09
$\delta = 0.75$	**1045.68**	$\ll$	1329.11	$\ll$	1412.36
$\delta = 0.9$	**950.68**	$\ll$	1061.13	$\gg$	1011.52
Mixed-Strategy Conjecture					
$\delta = 0.6$	1842.11	$\gg$	**1725.16**	$\ll$	1875.59
$\delta = 0.75$	1472.24	$\gg$	**1334.32**	$\ll$	1415.32
$\delta = 0.9$	1228.48	$\gg$	1073.86	$\gg$	**1023.88**
Behavior-Strategy Conjecture					
$\delta = 0.6$	2068.31	$\gg$	**1720.06**	$\approx$	1728.46
$\delta = 0.75$	1521.77	$\gg$	1262.84	$\gg$	**1202.79**
$\delta = 0.9$	1049.64	$\gg$	944.11	$\gg$	**732.88**

Note: Analysis based on simulated data sets comprising 50 subjects and 40 observations (past round 1) per subject, reporting the average ICL-BIC of the classes of fitted models. Here, $\gg, \ll$ indicate significance of differences (in Wilcoxon matched-pairs tests of the simulated ICL-BICs) at $\alpha = 0.01$ and $>, <$ indicate significance at $\alpha = 0.05$.

short supergames ($\delta = 0.6$) there appears to be a bias towards the random-switching model (mixed strategies). Our data set contains more observations for short supergames satisfying $\delta \leq 0.6$ than for long supergames satisfying $\delta \geq 0.9$, see Tables 14 and 15 in the appendix. Moreover, the average δ weighed by individual observations is around 0.73 in the first halves of sessions and 0.74 in the second halves, approximating the case where all conjectures are well-identified. Thus, in aggregate there may be a slight bias against semi-grim in the analysis, but aggregating across a large number of subject pools with $\delta = 0.75$ on average, our method seems suitable to reliably identify the correct model.

Table 7: Can we econometrically distinguish pure, mixed and behavior strategies? Robustness check for larger data

(a) Intermediate data set: 100 subjects

	Model fitted to the data				
	Finite Mixture		Random Switching		Semi-Grim
Pure-Strategy Conjecture					
$\delta = 0.6$	**2450.06**	$\ll$	3430.25	$\ll$	3746.59
$\delta = 0.75$	**2093.11**	$\ll$	2669.43	$\ll$	2841.81
$\delta = 0.9$	**1891.42**	$\ll$	2120.93	$\gg$	2042.6
Mixed-Strategy Conjecture					
$\delta = 0.6$	3693.23	$\gg$	**3454.03**	$\ll$	3762.16
$\delta = 0.75$	2948.44	$\gg$	**2673.63**	$\ll$	2835.13
$\delta = 0.9$	2457.96	$\gg$	2146.63	$\gg$	**2058.91**
Behavior-Strategy Conjecture					
$\delta = 0.6$	4132.82	$\gg$	**3436.66**	$\ll$	3463.7
$\delta = 0.75$	3047.1	$\gg$	2521.5	$\gg$	**2406.99**
$\delta = 0.9$	2105.78	$\gg$	1888.58	$\gg$	**1468.14**

(b) Large data set: around 200 subjects

	Model fitted to the data				
	Finite Mixture		Random Switching		Semi-Grim
Pure-Strategy Conjecture					
$\delta = 0.6$	**4908.54**	$\ll$	6870.11	$\ll$	7498.08
$\delta = 0.75$	**4175.89**	$\ll$	5329.73	$\ll$	5675.49
$\delta = 0.9$	**3791.25**	$\ll$	4252.75	$\gg$	4103.17
Mixed-Strategy Conjecture					
$\delta = 0.6$	7385.62	$\gg$	**6909.42**	$\ll$	7521.3
$\delta = 0.75$	5923.99	$\gg$	**5371.1**	$\ll$	5693.22
$\delta = 0.9$	4926.83	$\gg$	4298.72	$\gg$	**4120.77**
Behavior-Strategy Conjecture					
$\delta = 0.6$	8266.47	$\gg$	**6877.24**	$\ll$	6927.44
$\delta = 0.75$	6092.37	$\gg$	5037.9	$\gg$	**4805.57**
$\delta = 0.9$	4220.99	$\gg$	3787.16	$\gg$	**2938.41**

Note: Analysis based on simulated data sets comprising 50 subjects and 40 observations (past round 1) per subject, reporting the average ICL-BIC of the classes of fitted models. Here, $\gg, \ll$ indicate significance of differences (in Wilcoxon matched-pairs tests of the simulated ICL-BICs) at $\alpha = 0.01$ and $>, <$ indicate significance at $\alpha = 0.05$.

B Robustness check: Memory-2

We investigate memory length using a data mining approach similar to above. To this end, we extend the set of pure strategies to capture possible interdependence of actions with choices in $t-2$ and determine the best-fitting specification for each treatment. We then evaluate these best fitting specifications, treatment by treatment, against the above memory-1 model AD+SG, i.e. against the conjecture that all cooperating subjects homogeneously play a simple behavior strategy.

Specifically, we allow for two alternative approaches of extending the set of memory-1 strategies to memory-2. One approach follows Fudenberg et al. (2012), who introduced lenient and resilient variants of the pure memory-1 strategies, i.e., strategies that punish only after the second deviation or that punish for two rounds instead of one, respectively. Let us note that such variations in punishment behavior also follow if punishment is random as in memory-1 behavior strategies, which were not considered by Fudenberg et al. (2012). This first approach is applicable in particular to generalize pure memory-1 strategies, by providing a specific list of memory-2 generalizations. The other approach is novel and more generally allows the cooperation probabilities in round t to depend on the behavior of one or both players in $t-2$. Here, we allow for three different specifications: cooperation probabilities may be a function of the opponent's choice in $t-2$ (*TFT-Scheme*), a function of whether both players cooperated in $t-2$ or not (*Grim-Scheme*), or a function of the entire choice profile in $t-2$ (*General scheme*). This approach is parametric and suitable in particular to extend generalized pure strategies of type II (or, behavior strategies) from memory-1 to memory-2. As indicated, we set up this deliberately large number of ways to model memory-2 only to post-hoc pick the best of them for an evaluation against the memory-1 semi-grim specification.

Table 8 summarizes the results. First, we mine for mixtures of pure strategies, based on the list of 10 strategies[21] of Fudenberg et al. (2012). Given the above results, we assume that subjects do not switch strategies within half-sessions, as this comes without loss of descriptive adequacy for experienced subjects and only little loss for inexperienced subjects (for whom, however, memory-2 will turn out to be of negligible relevance). For each treatment, we determine the most adequate combination of strategies from a list of five possible combinations of Fudenberg et al.'s strategies, thus providing a selection of the best of 5^{32} models overall. The resulting model (Column "Best Pure M1&M2" in Table 8) fits highly significantly worse than the selection of pure and generalized-pure strategies with memory-1 defined above ("M1" in Table 8). We may therefore discard the possibility that subjects play pure strategies (with noise) of either memory-1 or memory-2, in favor of the possibility that they play generalized-pure strategies allowing for non-trivial randomization in at least one state.

Second, we take the above memory-1 model ("M1" in Table 8) as our benchmark and ask if equipping the pure or generalized pure strategies of type II with memory-2 improves goodness-of-fit. Again, we do so treatment by treatment. That is, for each treatment, we

[21] These strategies are TFT, Grim, AD, Grim2, TF2T, T2, 2TFT, 2PTFT as defined in Fudenberg et al. (2012) and also in Table 12 in the Online Appendix.

Table 8: **Memory-1 or Memory-2, and semi-grim, pure or generalized pure?** Strategy mixtures are estimated treatment-by-treatment. The resulting ICL-BICs are pooled within experiments and overall (less is better, relation signs point to better models)

	Memory-2 Generalizations of Semi-Grim + AD							Best Mixtures of Generalized Pure Strategies						Best Pure
	M2"General"		M2"TFT"		M2"Grim"		AD+SG	M1+M2"TFT"		M1+M2"Grim"		M1		M1 & M2
Specification														
# Models evaluated	1		1		1		1	22^{32}		22^{32}		13^{32}		5^{32}
# Pars estimated (by treatment)	12		6		6		5	160		160		80		32
# Parameters accounted for	12		6		6		5	6–15		6–15		6–10		3–8
First halves per session														
Aoyagi and Frechette (2009)	756.04	$\approx$	764.13	$\approx$	749.99	$\approx$	781.86	$\approx$	756.95	$\approx$	756.95	$\approx$	756.95	$\ll$	884.86
Blonski et al. (2011)	1244.76	$\gg$	1121.17	$\approx$	1120.87	$\gg$	1069.28	$\approx$	1069.56	$\approx$	1069.56	$\approx$	1069.58	$\approx$	1105.98
Bruttel and Kamecke (2012)	807.47	$\approx$	802.89	$\approx$	804.16	$\approx$	800.12	$\approx$	817.89	$\approx$	817.89	$\approx$	817.89	$\approx$	839.97
Dal Bó (2005)	660.68	$>$	641.34	$\approx$	642.26	$\approx$	629.17	$\approx$	635.04	$\approx$	635.04	$\approx$	635.04	$\approx$	653.05
Dal Bó and Fréchette (2011)	6671.28	$\approx$	6616.44	$\approx$	6604.7	$\approx$	6597.93	$\ll$	6904.79	$\approx$	6904.79	$\approx$	6904.79	$\ll$	7391.89
Dal Bó and Fréchette (2015)	8068.37	$\approx$	8028.83	$\approx$	8031.59	$\approx$	8017.59	$\ll$	8423.8	$\approx$	8431.51	$\approx$	8434.93	$\ll$	8893.78
Dreber et al. (2008)	805.74	$>$	785.48	$\approx$	785.6	$\approx$	782.37	$\approx$	787.71	$\approx$	787.71	$\approx$	787.71	$<$	863.47
Duffy and Ochs (2009)	1361.84	$\approx$	1377.17	$\approx$	1369.86	$\approx$	1372.97	$\approx$	1395.4	$\approx$	1395.4	$\approx$	1395.4	$\approx$	1426.34
Fréchette and Yuksel (2017)	305.9	$\approx$	299.72	$\approx$	296.93	$\approx$	299.62	$\approx$	300.87	$\approx$	300.87	$\approx$	300.87	$\approx$	317.35
Fudenberg et al. (2012)	387.8	$\approx$	379.84	$\approx$	378.07	$\approx$	381.01	$<$	432.32	$\approx$	432.32	$\approx$	432.32	$\approx$	463.4
Kagel and Schley (2013)	2542.02	$\approx$	2556.45	$\approx$	2552.09	$\approx$	2561.76	$\approx$	2679.23	$\approx$	2685.4	$\approx$	2685.4	$\approx$	2730.66
Sherstyuk et al. (2013)	1311.64	$\approx$	1307.45	$\approx$	1303.94	$\approx$	1303.8	$\approx$	1322.6	$\approx$	1322.6	$\approx$	1322.6	$<$	1398.69
Pooled	**25434.21**	$\gg$	**24972.71**	$\approx$	**24931.86**	$\approx$	**24779.85**	$\ll$	**25750.84**	$\approx$	**25757.44**	$\approx$	**25758.38**	$\ll$	**27115.39**
Second halves per session														
Aoyagi and Frechette (2009)	415.47	$\approx$	421.18	$>$	409.19	$\approx$	423.68	$\approx$	416.51	$\approx$	416.51	$\approx$	416.51	$\ll$	540.47
Blonski et al. (2011)	1518.54	$\gg$	1395.94	$\approx$	1393.41	$\gg$	1346.79	$\approx$	1398.5	$\approx$	1398.5	$\approx$	1398.5	$<$	1564.48
Bruttel and Kamecke (2012)	536.19	$\approx$	532.08	$\approx$	529.47	$\approx$	536.77	$\approx$	538.17	$\approx$	538.17	$\approx$	538.17	$\approx$	567.99
Dal Bó (2005)	727.25	$\approx$	710.88	$\approx$	708.32	$\approx$	699.05	$\approx$	726.04	$\approx$	731.81	$\approx$	732.27	$\approx$	741.2
Dal Bó and Fréchette (2011)	5201.05	$\approx$	5137.82	$\approx$	5132.96	$\approx$	5128.69	$\approx$	5195.88	$\approx$	5195.88	$\approx$	5195.88	$\ll$	5960.78
Dal Bó and Fréchette (2015)	7840.87	$\approx$	7829.51	$\approx$	7808.63	$\approx$	7825.98	$\ll$	8172.63	$\approx$	8177.46	$\approx$	8177.46	$\ll$	9143.98
Dreber et al. (2008)	597.17	$\approx$	580.63	$\approx$	570.33	$\approx$	589.84	$\approx$	618.5	$\approx$	618.89	$\approx$	619.9	$\approx$	648.55
Duffy and Ochs (2009)	1706.1	$\approx$	1753.41	$\approx$	1719.86	$\approx$	1761.6	$\approx$	1857.06	$\approx$	1876.72	$\approx$	1883.52	$\approx$	2003.41
Fréchette and Yuksel (2017)	422.32	$\approx$	424.41	$\approx$	419.44	$\approx$	423.34	$\approx$	433.18	$\approx$	433.18	$\approx$	433.18	$<$	464.23
Fudenberg et al. (2012)	452.64	$\approx$	450.08	$\approx$	447.25	$\approx$	452.6	$<$	484.5	$\approx$	477.91	$\approx$	514.87	$\approx$	534.47
Kagel and Schley (2013)	1782.43	$\approx$	1777.83	$\approx$	1773.55	$\approx$	1775.62	$\approx$	1751.81	$\approx$	1751.81	$\approx$	1751.81	$\approx$	1830.26
Sherstyuk et al. (2013)	959.21	$\approx$	952.56	$\approx$	949.46	$\approx$	951.34	$\approx$	955.73	$\approx$	955.73	$\approx$	955.73	$\approx$	1023.43
Pooled	**22669.91**	$\gg$	**22258.14**	$\approx$	**22153.69**	$\approx$	**22097.67**	$\ll$	**22811.34**	$\approx$	**22828.13**	$\approx$	**22848.49**	$\ll$	**25177.57**

Note: Results treatment-by-treatment are in the appendix. The main body contains ICL-BICs aggregated at paper level. Relation signs and p-values are exactly as above, see Table 3. "M2" ("M1") denotes strategies, whose actions may depend on actions in $t-2$ and $t-1$ ($t-1$ only). The supplements "General", "TFT", "Grim" indicate whether parameters of behavior strategies may depend on: all four possible histories in $t-2$ (M2 "General"), whether the opponent cooperated in $t-2$ (M2 "TFT"), or whether there was joint cooperation in $t-2$ (M2 "Grim"). Pure M2 strategies do not have such free parameters. Columns 1-3 contain one memory-2 version of semi-grim each. Column 4 is memory-1 semi-grim. Columns 5-7 are memory-2 and memory-1 versions of generalized prototypical strategies. The last column contains the best fitting combinations of a set of pure memory-1 and memory-2 strategies from the literature (TFT, Grim, AD, Grim2, TF2T, T2, 2TFT, 2PTFT) for definitions see Table 12 in the Online Appendix.

take the best-fitting of the 13 memory-1 models discussed above, the best of the five pure memory-2 strategy combinations following Fudenberg et al. (2012), and the best of the four generalized pure strategies (type II) after allowing for them to be of memory-2 following either the TFT scheme or the Grim scheme, and then take the best of these 22 models overall. The results are provided in the columns *M1+M2"TFT"* and *M1+M2"Grim"* of Table 8: After allowing for generalized pure strategies as done here, allowing for memory-2 has virtually zero impact for inexperienced subjects and some but only insignificant impact for experienced subjects.[22] This indicates that the appearance of memory-2 is indistinguishable from the parametrically simpler notion of randomization as in generalized-pure strategies. Further, all of these data-mined models still fit significantly worse than the simple AD+SG that stays free from post-hoc modeling choices (column 4 of Table 8). Considering that the best of 22^{32} models, comprising all of the key ideas expressed in behavioral analyses of repeated games, does not improve on this single model now strongly indicates that subjects actually play behavior strategies.

Third, we evaluate whether these behavior strategies possibly have memory-2. That is, we compare the simple AD+SG memory-1 version with the three generalizations to memory-2 introduced above. The TFT-scheme allows the cooperation probabilities to be functions of the opponent's action in $t-2$, the Grim-scheme allows them to be functions of whether both subjects cooperated in $t-2$, and the General scheme of all four possible states in $t-2$. The results are report in the three left-most columns of Table 8 and appear clear-cut: None of the memory-2 extensions improves on describing behavior by the simple memory-1 semi-grim strategy. Indeed, the finer the memory-2 ramifications, the worse the model adequacy (after accounting for the additional degrees of freedom). These results are additionally compatible with a result of Breitmoser (2015) who verified the Markov assumption by testing whether subjects systematically deviate from memory-1 strategies after particular histories in memory-2. We summarize these observations as follows.

Result 6 (Memory-2)**.** *Model adequacy does not improve by equipping subjects with memory-2, neither for (generalizations of) pure strategies nor for semi-grim.*

C Further details and results on the structural analysis of preferences and beliefs

Our objective is to examine to what extent received models are compatible with the observation that the (sub-)population of cooperating subjects consists of two components, cautious cooperators and strong cooperators, who play the mildly treatment-dependent strategies estimated above. For clarity, the estimated strategies are listed in Table 9. In the structural analysis, we do not include defecting subjects, as their behavior is easily rationalizable across treatments. Further, we do not seek to model the relative shares of cautious cooperators and

[22] One reason for the good performance of generalized memory-1 strategies compared to the generalized memory-2 strategies is that allowing for first round randomization seems essential. However, when we abstract from first rounds, as done in an earlier draft (available from the authors), we obtained a similarly bad fit for generalized memory-2 strategies compared to generalized memory-1 strategies (both type-II generalization).

strong cooperators, since the shares seem closely related to existing predictors of cooperation. The actual strategies leave us with the remaining part of our original research question to understand behavior in the repeated PD: How can we rationalize this behavior?

In order to introduce the requisite notation in a general setting, let us consider a player using strategy σ (as above, a mapping from memory-1 states to probabilities of cooperation) with initially arbitrary beliefs about the possibly types of opponents. The set of opponents' types is K, and opponents of type $k \in K$ play strategy τ_k. The prior belief of facing opponent type $k \in K$ is denoted ρ_k, and given history h, the posterior belief that the opponent is of type $k \in K$ is denoted as $\Pr(k|h)$. Obviously, this posterior is a function of the prior belief (K, ρ, τ), which we shall make explicit in the notation below. Define $\tau = \{\tau_k\}_k$. The expected payoff of playing $a \in \{c,d\}$ after history h, over the present and all subsequent rounds of the indefinitely repeated game, given one's own continuation strategy σ and the opponent's strategy τ_k, is denoted as $\Pi(a|h,\sigma,\tau_k)$. Holding the belief $\Pr(k|h)$ fixed, the expected payoff of playing $a \in \{c,d\}$, given σ and τ, can be written as

$$\Pi^0(a|h,\sigma,K,\rho,\tau) = \sum_{k \in K} \Pr(k|h,K,\rho,\tau) \cdot \Pi(a|h,\sigma,\tau_k). \tag{14}$$

Assuming logistic errors and precision $\lambda \geq 0$, the probability of observing action $a \in \{c,d\}$ in state ω is thus

$$\Pr(a,\omega|\sigma,K,\rho,\tau) = \frac{\exp\{\lambda \cdot \Pi^0(a|\omega,\sigma,K,\rho,\tau)\}}{\exp\{\lambda \cdot \Pi^0(c|\omega,\sigma,K,\rho,\tau)\} + \exp\{\lambda \cdot \Pi^0(d|\omega,\sigma,K,\rho,\tau)\}}. \tag{15}$$

Now, let the estimated population be described by the two types $(\hat{\rho}_{\text{cautious}}, \hat{\sigma}_{\text{cautious}})$ and $(\hat{\rho}_{\text{strong}}, \hat{\sigma}_{\text{strong}})$, and let the underlying data set consist of $n(a,\omega)$ observations of action a in state ω. Allowing cautious and strong cooperators to hold different beliefs, the log-likelihood of a belief model (ρ, τ, K) with respect is

$$\begin{aligned} LL(\rho,\tau,K) = \hat{\rho}_{\text{cautious}} \sum_{\omega} \sum_{a \in \{c,d\}} n(a,\omega) \cdot \log \Pr(a,\omega|\hat{\sigma}_{\text{cautious}}, K_c, \rho_c, \tau_c) \\ + \hat{\rho}_{\text{strong}} \sum_{\omega} \sum_{a \in \{c,d\}} n(a,\omega) \cdot \log \Pr(a,\omega|\hat{\sigma}_{\text{strong}}, K_s, \rho_s, \tau_s). \end{aligned}$$

We maximize this log-likelihood using the same algorithms as above (first NEWUOA and Newton-Raphson), where the free parameters are those of the utility models described below.

Modeling prior beliefs We consider the following three types of prior beliefs. The "naive" prior assumes that opponents are homogeneous and play an average strategy, the "correct" prior assumes that all three types of opponents exist, and the "consensus" prior assumes that opponents are of the same type as oneself. Using $(\tilde{K}, \tilde{\rho}, \tilde{\tau})$ to denote the true type distributions, the beliefs (K, ρ, τ) of the two playe types (c,s), i.e. cautious and strong cooperators,

Table 9: Estimated cooperation probabilities and shares of the identified player types

		Defectors		Cautious Coop.		Strong Coop.		Continuation of Coop.		
Experiment/Treatment	$\delta - \delta^*$	Share	ε	Share	$\sigma_\emptyset$	Share	$\sigma_\emptyset$	σ_{cc}	$\sigma_{cd/dc}$	σ_{dd}
First halves per session										
DF11–6	-0.32	0.487	0.016	0.47	0.181	0.05	0.887	0.922	0.398	0.078
DF15–4	-0.32	0.591	0.026	0.39	0.263	0.02	0.99	0.895	0.356	0.105
BOS11–9	-0.3	0.366	0	0.59	0.304	0.05	0.99	0.946	0.201	0.054
BOS11–15	-0.15	0.839	0.008	0.08	0.196	0.08	0.197	0.999	0.224	0.001
DF11–7	-0.11	0.308	0.018	0.4	0.123	0.29	0.43	0.894	0.324	0.106
DF11–22	-0.07	0.316	0.013	0.42	0.212	0.27	0.568	0.916	0.383	0.084
DF15–20	-0.07	0.276	0.01	0.61	0.247	0.11	0.882	0.921	0.322	0.079
BOS11–14	-0.05	0.189	0.203	0.73	0.069	0.08	0.478	0.991	0.123	0.009
BOS11–26	-0.05	0.174	0.222	0.62	0.112	0.21	0.831	0.984	0.171	0.016
DRFN08–10	-0.05	0.188	0.036	0.62	0.438	0.19	0.965	0.948	0.178	0.052
BOS11–30	0.07	0.648	0.062	0.21	0.484	0.14	0.99	1	0	0
BOS11–31	0.07	0.33	0.027	0.27	0.256	0.4	0.895	0.977	0.512	0.023
BOS11–16	0.08	0.051	0.5	0.49	0.251	0.46	0.898	0.95	0.178	0.05
BOS11–27	0.08	0.502	0.01	0.35	0.373	0.15	0.99	0.887	0.448	0.113
D05–18	0.08	0.096	0.045	0.36	0.144	0.55	0.782	0.86	0.286	0.14
D05–19	0.08	0.233	0.01	0.44	0.289	0.33	0.956	0.914	0.335	0.086
DF15–33	0.08	0.279	0.025	0.55	0.291	0.17	0.898	0.929	0.368	0.071
DRFN08–11	0.08	0.271	0.052	0.39	0.468	0.34	0.908	0.931	0.329	0.069
DF11–8	0.11	0.251	0.01	0.32	0.204	0.43	0.699	0.906	0.419	0.094
DF15–5	0.11	0.296	0.095	0.31	0.458	0.39	0.947	0.933	0.309	0.067
BK12–28	0.13	0.143	0.077	0.47	0.262	0.39	0.867	0.916	0.289	0.084
DF15–35	0.13	0.169	0.167	0.3	0.076	0.54	0.833	0.972	0.417	0.028
DF11–23	0.14	0.21	0.08	0.25	0.288	0.54	0.817	0.951	0.458	0.049
KS13–12	0.15	0.224	0.022	0.31	0.431	0.47	0.911	0.932	0.335	0.068
BOS11–17	0.18	0.569	0.298	0.14	0.036	0.29	0.75	1	0.383	0
STS13–13	0.19	0.152	0.074	0.33	0.275	0.52	0.892	0.919	0.409	0.081
DO09–32	0.23	0.22	0.1	0.31	0.283	0.47	0.905	0.901	0.373	0.099
FY17–25	0.31	0.119	0.01	0.32	0.581	0.56	0.984	0.926	0.245	0.074
DF11–24	0.36	0.112	0.189	0.37	0.597	0.51	0.948	0.949	0.356	0.051
DF15–21	0.36	0.186	0.088	0.24	0.407	0.58	0.926	0.942	0.467	0.058
FRD12–29	0.48	0.062	0.023	0.3	0.417	0.63	0.983	0.97	0.469	0.03
AF09–34	0.59	0.15	0.5	0.53	0.648	0.32	0.988	0.911	0.41	0.089
Second halves per session										
DF11–6	-0.32	0.687	0.01	0.27	0.093	0.05	0.643	0.937	0.553	0.063
DF15–4	-0.32	0.635	0.009	0.28	0.102	0.08	0.72	0.94	0.234	0.06
BOS11–9	-0.3	0.166	0.074	0.53	0.01	0.3	0.718	0.999	0.129	0.001
BOS11–15	-0.15	0.008	0.007	0.9	0.001	0.09	0.001	0.998	0.001	0.002
DF11–7	-0.11	0.399	0.009	0.41	0.178	0.19	0.615	0.864	0.473	0.136
DF11–22	-0.07	0.313	0.01	0.46	0.158	0.23	0.818	0.963	0.465	0.037
DF15–20	-0.07	0.392	0.01	0.4	0.179	0.21	0.856	0.943	0.42	0.057
BOS11–14	-0.05	0.002	0.01	0.95	0	0.05	0.491	0.987	0.3	0.013
BOS11–26	-0.05	0.43	0.008	0.39	0.365	0.18	0.748	0.935	0.291	0.065
DRFN08–10	-0.05	0.455	0.01	0.4	0.342	0.14	0.908	0.968	0.252	0.032
BOS11–30	0.07	0.339	0.01	0.57	0.356	0.09	0.99	0.963	0.201	0.037
BOS11–31	0.07	0.449	0.017	0.23	0.139	0.33	0.88	0.979	0.484	0.021
BOS11–16	0.08	0.11	0.371	0.43	0.271	0.46	0.977	0.966	0.208	0.034
BOS11–27	0.08	0.355	0.01	0.34	0.109	0.3	0.887	0.951	0.495	0.049
D05–18	0.08	0.071	0.009	0.38	0.048	0.55	0.83	0.878	0.396	0.122
D05–19	0.08	0.21	0.018	0.16	0.051	0.63	0.825	0.947	0.295	0.053
DF15–33	0.08	0.219	0.01	0.38	0.159	0.4	0.841	0.964	0.476	0.036
DRFN08–11	0.08	0.091	0.01	0.3	0.309	0.61	0.92	0.951	0.327	0.049
DF11–8	0.11	0.37	0.01	0.24	0.231	0.39	0.896	0.971	0.446	0.029
DF15–5	0.11	0.291	0.021	0.3	0.345	0.41	0.948	0.963	0.322	0.037
BK12–28	0.13	0.236	0.015	0.42	0.275	0.34	0.969	0.948	0.323	0.052
DF15–35	0.13	0.156	0.01	0.31	0.127	0.53	0.93	0.967	0.51	0.033
DF11–23	0.14	0.079	0.01	0.16	0.151	0.76	0.967	0.956	0.508	0.044
KS13–12	0.15	0.165	0.01	0.12	0.175	0.71	0.954	0.964	0.359	0.036
BOS11–17	0.18	0.311	0.009	0.53	0.504	0.16	0.908	0.95	0.256	0.05
STS13–13	0.19	0.125	0.021	0.25	0.237	0.63	0.925	0.953	0.55	0.047
DO09–32	0.23	0.047	0.01	0.3	0.173	0.66	0.953	0.954	0.392	0.046
FY17–25	0.31	0.139	0.024	0.14	0.514	0.73	0.956	0.957	0.352	0.043
DF11–24	0.36	0	0.053	0.11	0.647	0.89	0.99	0.98	0.334	0.02
DF15–21	0.36	0.089	0.01	0.16	0.337	0.75	0.958	0.965	0.373	0.035
FRD12–29	0.48	0.083	0.06	0.09	0.119	0.83	0.967	0.965	0.536	0.035
AF09–34	0.59	0.133	0.498	0.05	0.259	0.81	0.984	0.968	0.461	0.032

are formally defined as follows.

$$\begin{array}{rlll}
\text{Naive:} & K_c = K_s = \{A\} & \rho_A = 1 & \tau_A = \sum_{k \in \tilde{K}} \tilde{\rho}_k \tilde{\tau}_k \\
\text{Correct:} & K_c = K_s = \tilde{K} & \rho_k = \tilde{\rho}_k & \tau_k = \tilde{\tau}_k \ \forall k \in \tilde{K} \\
\text{Consensus:} & K_c = \{c\} & \rho_c = 1 & \tau_c = \tilde{\tau}_{\text{cautious}} \\
 & K_s = \{s\} & \rho_s = 1 & \tau_s = \tilde{\tau}_{\text{strong}}
\end{array}$$

Bayesian updating of beliefs Players with correct beliefs understand that subjects are not homogeneous and therefore update their beliefs given their observations. Using Bayes' rule, the posterior belief after history h is

$$\Pr(k|h,K,\rho,\tau) = \frac{\rho_k \Pr(h|\sigma,\tau_k)}{\sum_{k' \in K} \rho_{k'} \Pr(h|\sigma,\tau_{k'})},$$

where $\Pr(h|\sigma,\tau_k)$, with $k \in \{A,B\}$, denotes the probability that history h is reached if the own strategy is σ and the opponent's strategy is τ_k. As estimated above, subjects in experiments seem to condition their actions on memory-1 Markov states, as opposed to more complex subsets of the history or even entire histories. This form of bounded rationality (i.e., imperfect recall) needs to be acknowledged, but can be expressed straightforwardly also in belief formation. Given the own strategy σ and the two opponent types' strategies τ_k, the **memory-1 posterior** that the opponent's type is $k \in K$ given memory-1 state ω is

$$\Pr(k|\omega,K,\rho,\tau) = \frac{\rho_k \sum_{h \in H(\omega)} \Pr(h|\sigma,\tau_k)}{\sum_{k' \in K} \rho_{k'} \sum_{h \in H(\omega)} \Pr(h|\sigma,\tau_{k'})},$$

where $H(\omega)$ is the set of histories leading to the memory-1 state ω.

Interdependent preferences As discussed above, we also examine to what extent received models of interdependent preferences allow us to capture behavior—having observed that pure payoff concerns are inevitably insufficient to capture behavior across treatments. The extension of the above definitions from expected payoffs to expected utilities is straightforward using the following definitions of stage game utilities. To begin with, all models of interdependent preferences are defined such that they allow for two free parameters. In **altruism**, we allow for the payoff of the other player to be relevant, and to obtain two free parameters as in other models, the other payoff's weight is allowed to depend on the relation of the own payoff to any reference point in $[0,1]$. In **inequity aversion**, we use a standard implementation of Fehr-Schmidt preferences. In **conditional cooperation**, we allow the utility to express an aversion against unilateral cooperation and unilateral defection (i.e. a preference for matching the opponent's action). In **generalized fairness**, we generalize the parameter-free fairness concerns fo Rabin (1993) to contain two free parameters just like the other models.

$$
\begin{aligned}
\text{Altruism:} \quad & u(\pi_1,\pi_2,a_1,a_2) = \pi_1 + I_{\pi_1 \geq 0.5} \cdot \alpha\pi_2 + I_{\pi_1 < 0.5} \cdot \beta\pi_2 \\
\text{Inequity aversion:} \quad & u(\pi_1,\pi_2,a_1,a_2) = \pi_1 - I_{\pi_1 \geq \pi_2} \cdot \alpha\pi_2 - I_{\pi_1 < \pi_2} \cdot \beta\pi_2 \\
\text{Cond. cooperation:} \quad & u(\pi_1,\pi_2,a_1,a_2) = \pi_1 - I_{a_1=d \wedge a_2=c} \cdot \alpha g - I_{a_1=c \wedge a_2=d} \cdot \beta l
\end{aligned}
$$

Our definition of **generalized fairness concerns** requires additional notation. Recall that Rabin (1993) definitions imply that in a one-shot PD with probabilities of cooperation $(s_1, s_2) \in [0,1]^2$, player i's utility is

$$
\begin{aligned}
U_i(s_i,s_j) &= \pi_i(s_i,s_j) + \overline{f}_j(s_j,s_i) \cdot f_i(s_i,s_j) \\
&= s_j \cdot (1+g) - s_i \cdot l - s_i s_j \cdot (g-l) + (s_j - 1/2)(s_i - 1/2).
\end{aligned}
$$

We generalize this towards

$$
U_i(s_i,s_j) = s_j \cdot (1+g) - s_i \cdot l - s_i s_j \cdot (g-l) + \alpha(s_j - \beta)(s_i - \beta) \tag{16}
$$

and as for the implicit stage game payoffs, this implies

$$
\begin{aligned}
U_i(1,1) &= 1 + \alpha(1-\beta)^2 = 1 + \alpha(1 - 2\beta + \beta^2) &&\hat{=} 1 + \alpha(1-2\beta) \\
U_i(1,0) &= -l - \alpha\beta(1-\beta) = -l - \alpha\beta + \alpha\beta^2 &&\hat{=} -l - \alpha\beta \\
U_i(0,1) &= 1 + g - \alpha\beta(1-\beta) = 1 + g - \alpha\beta + \alpha\beta^2 &&\hat{=} 1 + g - \alpha\beta \\
U_i(0,0) &= \alpha\beta^2 &&\hat{=} 0,
\end{aligned}
$$

i.e. that the players play a constituent game resembling a "PD" with the parameters $l^* = \frac{l+\alpha\beta}{1+\alpha(1-2\beta)}$ and $g^* = \frac{1+g-\alpha\beta}{1+\alpha(1-2\beta)} - 1$.

Discounting We allow for the perceived discount factor $\tilde{\delta}$ to be a function of the true discount factor as in $\tilde{\delta} = \delta^x$. If $x = 1$, subjects correctly perceive the discount factor (or, break-up probability), for $x < 1$ they underestimate it, with the limiting case $x \to 0$ where they simply disregard the break-up probabibility and simply play the game as if it had an infinite time horizon (or, without impatience). In turn, if $x > 1$, subjects overestimate the break-up probability, and in the limiting case $x \to \infty$, subjects are myopic and play a sequence of one-shot games. In the estimation x is limited to 100 for viability.

Parametrization Overall, all models thus have up to three parameters, exponent X characterizing the perceived discount factor δ^X and (α, β) characterizing the extent of social preferences.

Benchmarks We provide results for the standard benchmark of *uniform randomization*, i.e. the goodness-of-fit of predicting 50-50 randomization in all states, and for the benchmark *clairvoyance* predicting the actually estimated probabilities of cooperation in all states. The

first one is a lower bound of the goodness-of-fit and the second one is an upper bound, and in relation to those we can estimate the extent to which behavior is explained by the various model components.

$$LL_{\text{Random}}(\rho,\tau,K) = \hat{\rho}_{\text{cautious}} \sum_{\omega} \sum_{a\in\{c,d\}} n(a,\omega)\cdot \log 1/2 + \hat{\rho}_{\text{strong}} \sum_{\omega} \sum_{a\in\{c,d\}} n(a,\omega)\cdot \log 1/2$$

$$\begin{aligned} LL_{\text{Clairvoyance}}(\rho,\tau,K) = {} & \hat{\rho}_{\text{cautious}} \sum_{\omega} \sum_{a\in\{c,d\}} n(a,\omega)\cdot \log \hat{\sigma}_{\text{cautious}}(a,\omega) \\ & + \hat{\rho}_{\text{strong}} \sum_{\omega} \sum_{a\in\{c,d\}} n(a,\omega)\cdot \log \hat{\sigma}_{\text{strong}}(a,\omega). \end{aligned}$$

BIC Instead of looking at the pure log-likelihoods, we evaluate models based on their Bayes information criteria $BIC = -LL + \#pars \cdot \log \#obs/2$, reported in Tables 5 in the paper, and Tables 10 and 11 in the appendix. As for the two benchmark models, whose specification remains the same across the three column sets, we evaluate the BIC using the numbers of parameters and observations for the models that are compared to the respective benchmark.This way, we obtain upper and lower bounds for the reported BIC.

Figure 5: Relation of observed and predicted probabilities of cooperation (second halves of sessions)

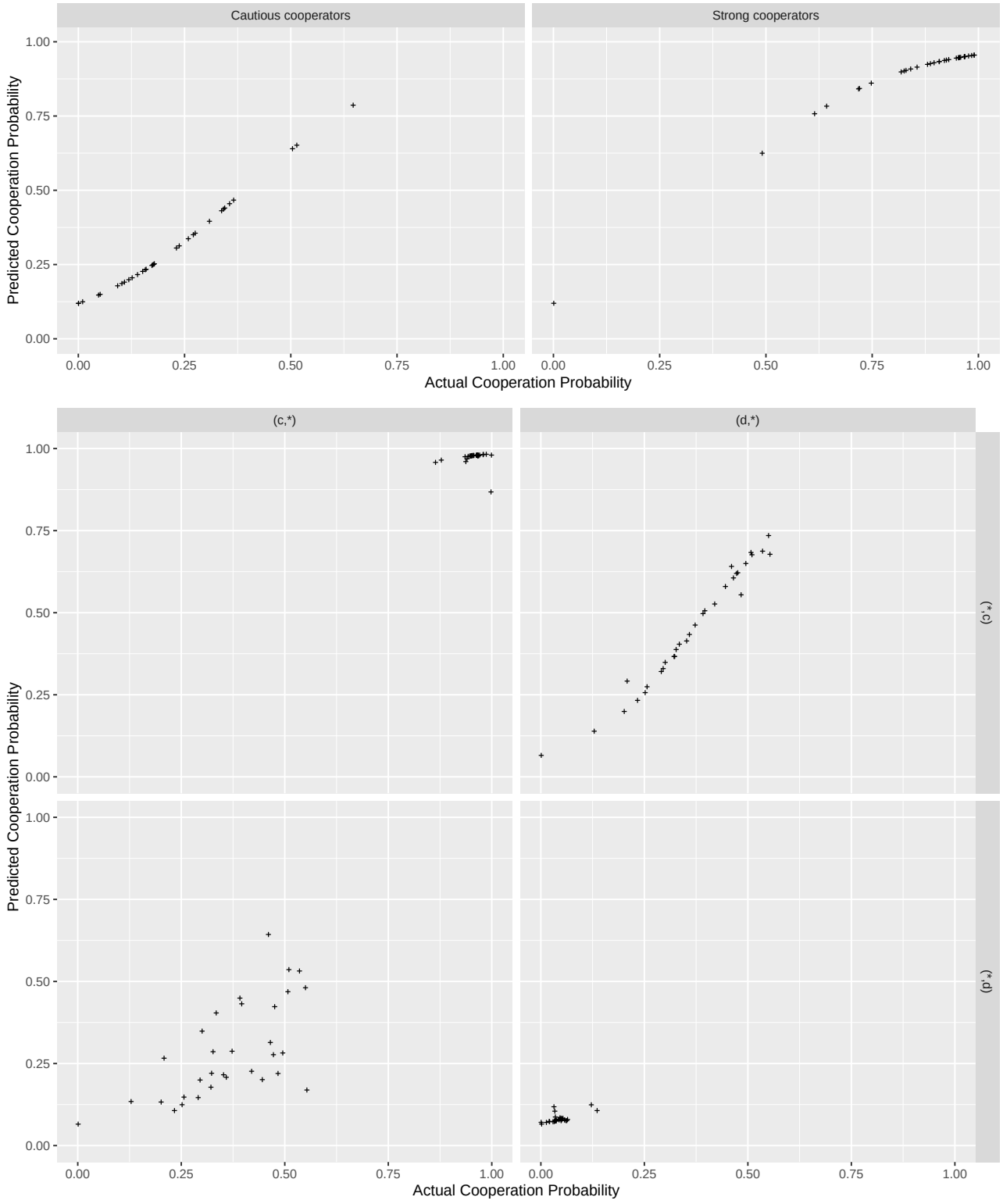
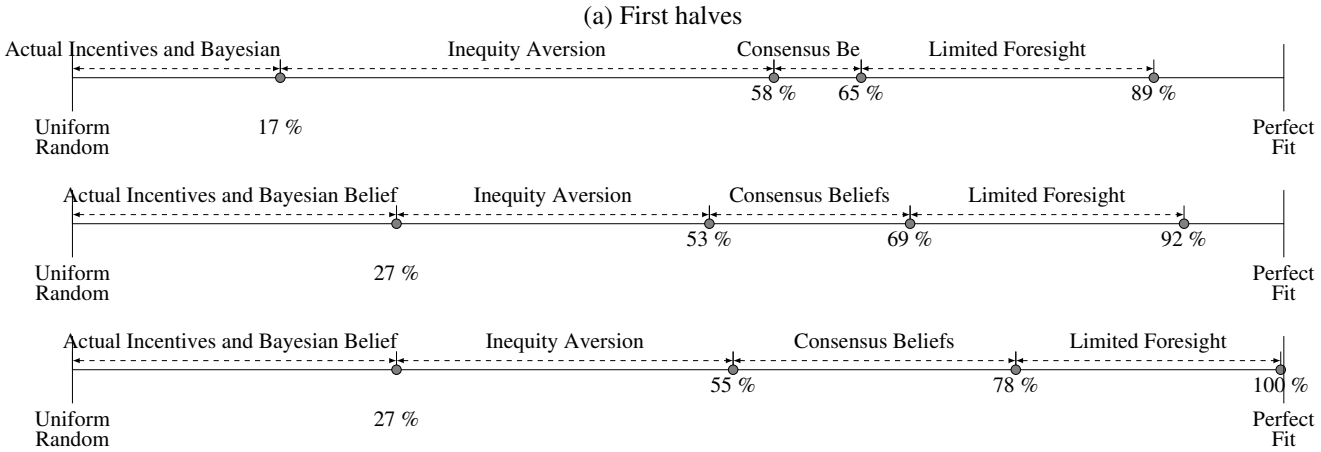


Figure 6: Decomposition of the structural model components
Top graph in each half: with constant preference parameters and constant variance of noise;
Middle graph in each half: with constant preference parameters and treatment-dependent variance of noise;
Bottom graph in each half: with treatment-dependent preference parameters and variance of noise



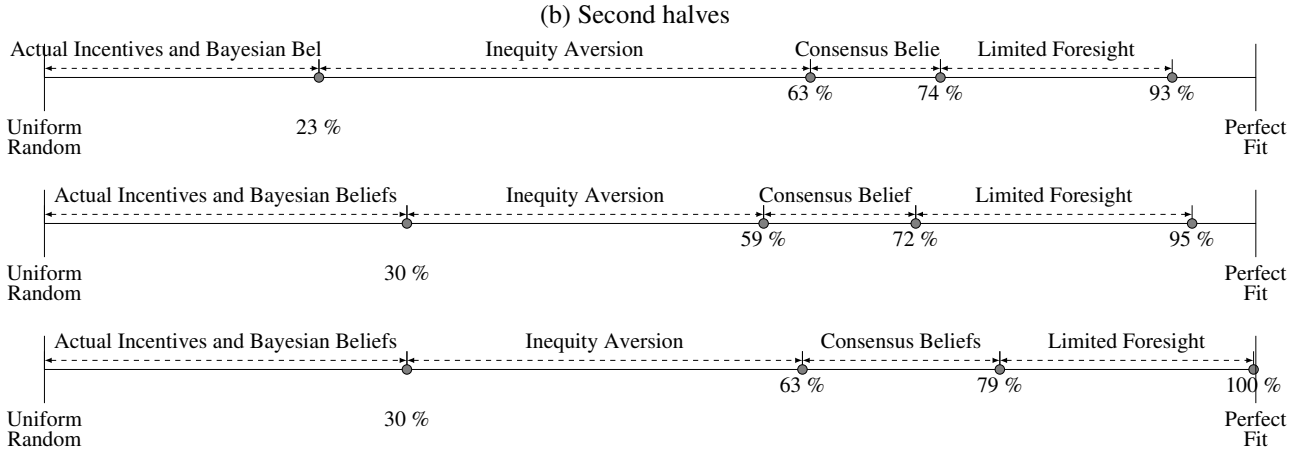


Table 10: Testing interdependence of preferences (both halves)

Model (free parameters)	Fit to pooled data				Fit to each treatment	
	Homogeneous variance		Heterogeneous variance			
	BIC	Estimates	*BIC*	Estimates	*BIC*	Average Estimates
Upper bound BIC (Clairvoyance)	45129.7		45361.8		46058	
Lower bound BIC (Uniform Random)	94380.8		94612.8		95309.1	
False Consensus Beliefs						
True supergame (g,l,δ), no free par $(-)$	83654.6	$(-,-,-)$	79175.3	$(-,-,-)$	79871.6	$(-,-,-)$
True stage game g,l, free $(\delta^X,-,-)$	83455	$(1.19,-,-)$	78282.6	$(1.35,-,-)$	74836.7	$(1.3,-,-)$
True δ, inequity aversion $(-,\alpha,\beta)$	61047.3	$(-,1.01,0.5)$	60561.3	$(-,1.28,0.6)$	55375.4	$(-,15.44,0.43)$
Inequity Aversion (δ^X,α,β)	49685.8	$(100,0.79,0.12)$	48920.5	$(19.85,0.75,0.1)$	46200.5	$(46.62,0.7,-0.05)$
Cond Cooperation (δ^X,α,β)	55923	$(100,1.48,-0.37)$	54269.2	$(5.96,1.6,-0.05)$	46599.6	$(20.26,2.01,-0.03)$
Altruism (δ^X,α,β)	53154.9	$(74.75,1.37,-0.27)$	51208.9	$(21.84,1.33,-0.22)$	46187.8	$(9.72,0.89,0.24)$
Gen Fairness Equilibrium (δ^X,α,β)	57075.3	$(7.4,38.23,0.22)$	54040.5	$(6.57,43.25,0.22)$	47108.2	$(9.12,33.11,0.02)$
Naive Beliefs						
True supergame (g,l,δ), no free par $(-)$	83743.8	$(-,-,-)$	81266.6	$(-,-,-)$	81962.8	$(-,-,-)$
True stage game g,l, free $(\delta^X,-,-)$	83437	$(1.14,-,-)$	80676.1	$(1.21,-,-)$	77696.3	$(3.62,-,-)$
True δ, inequity aversion $(-,\alpha,\beta)$	61994.4	$(-,-100,-3.67)$	62929.2	$(-,-100,-2.69)$	60135	$(-,-100,-2.36)$
Inequity Aversion (δ^X,α,β)	56552	$(100,26.11,1.09)$	56390.1	$(87.42,3.36,1.09)$	56398.5	$(49.69,14.98,0.82)$
Cond Cooperation (δ^X,α,β)	67945.9	$(100,30.4,0.22)$	63489.7	$(4.24,8.45,0.47)$	56398.5	$(100,20.26,-0.05)$
Altruism (δ^X,α,β)	59484.3	$(87.42,15.46,-0.96)$	58236.5	$(19.53,44.06,-0.8)$	56398.5	$(34.4,15.85,-0.59)$
Gen Fairness Equilibrium (δ^X,α,β)	59345.5	$(4.13,-6.15,0.53)$	57955.1	$(3.93,-6.09,0.54)$	56398.5	$(40.1,9.13,0.41)$
Bayesian Beliefs						
True supergame (g,l,δ), no free par $(-)$	83891.8	$(-,-,-)$	80092.5	$(-,-,-)$	80788.7	$(-,-,-)$
True stage game g,l, free $(\delta^X,-,-)$	83746	$(0.9,-,-)$	80091.6	$(1.01,-,-)$	77163.4	$(1.39,-,-)$
True δ, inequity aversion $(-,\alpha,\beta)$	62553.9	$(-,11.84,1.21)$	65804.6	$(-,2.23,0.86)$	65085.3	$(-,100,100)$
Inequity Aversion (δ^X,α,β)	56374.2	$(100,9.72,0.98)$	56319.6	$(87.42,11.68,0.95)$	56706	$(6.83,100,0.9)$
Cond Cooperation (δ^X,α,β)	68640.9	$(100,100,0.11)$	63818.6	$(3.69,14.65,0.33)$	56704.2	$(22.88,5.08,-0.22)$
Altruism (δ^X,α,β)	58022.9	$(100,-100,5.74)$	57485.5	$(11.32,-100,5.53)$	56704.8	$(20.47,100,-0.7)$
Gen Fairness Equilibrium (δ^X,α,β)	66459.5	$(5.83,17.42,0.21)$	63148	$(5.52,7.93,0.2)$	56704.1	$(9.02,81.49,-0.26)$

Table 11: Testing interdependence of preferences (first halves)

Model (free parameters)	Fit to pooled data: Homogeneous variance *BIC*	Fit to pooled data: Homogeneous variance Estimates	Fit to pooled data: Heterogeneous variance *BIC*	Fit to pooled data: Heterogeneous variance Estimates	Fit to each treatment *BIC*	Fit to each treatment Average Estimates
Upper bound BIC (Clairvoyance)	22650.9		22883		23579.2	
Lower bound BIC (Uniform Random)	42075.3		42307.3		43003.5	
False Consensus Beliefs						
True supergame (g,l,δ), no free par $(-)$	38037.9	$(-,-,-)$	36156.8	$(-,-,-)$	36853	$(-,-,-)$
True stage game g,l, free $(\delta^X,-,-)$	37892.6	$(1.28,-,-)$	35739.7	$(1.36,-,-)$	34543.1	$(0.4,-,-)$
True δ, inequity aversion $(-,\alpha,\beta)$	29429.8	$(-,1.02,0.49)$	28878	$(-,1.32,0.66)$	27879.8	$(-,11.37,0.54)$
Inequity Aversion (δ^X,α,β)	24738.5	$(100,0.81,0.14)$	24486.5	$(24.6,0.76,0.12)$	23632.4	$(57.13,0.61,-0.02)$
Cond Cooperation (δ^X,α,β)	27236.5	$(100,1.48,-0.32)$	26859.8	$(100,1.55,-0.32)$	23848.4	$(2.8,2.04,0)$
Altruism (δ^X,α,β)	26204.3	$(77.34,1.32,-0.27)$	25520.2	$(28.27,1.29,-0.23)$	23633.4	$(10.43,3.43,0.12)$
Gen Fairness Equilibrium (δ^X,α,β)	27087.2	$(8.91,2.95,0.22)$	25933.4	$(6.24,3.07,0.24)$	24047.4	$(5.89,6.9,0.04)$
Naive Beliefs						
True supergame (g,l,δ), no free par $(-)$	39017.8	$(-,-,-)$	37922	$(-,-,-)$	38618.3	$(-,-,-)$
True stage game g,l, free $(\delta^X,-,-)$	38910.2	$(1.14,-,-)$	37674	$(1.22,-,-)$	36580.1	$(0.6,-,-)$
True δ, inequity aversion $(-,\alpha,\beta)$	30817.9	$(-,-87.62,-3.64)$	31187.3	$(-,-87.77,-2.91)$	30333.6	$(-,-69.79,-2.46)$
Inequity Aversion (δ^X,α,β)	28610.8	$(100,8.31,1.04)$	28622.1	$(87.42,4.61,1.04)$	28969.7	$(55.54,100,0.72)$
Cond Cooperation (δ^X,α,β)	33552.7	$(100,26.39,0.31)$	31888	$(3.34,4.55,0.57)$	28967.9	$(62.27,9.96,-0.08)$
Altruism (δ^X,α,β)	29758.9	$(100,17.51,-0.95)$	29435.6	$(22.6,46.57,-0.77)$	28969.7	$(40.36,34.44,-0.47)$
Gen Fairness Equilibrium (δ^X,α,β)	29944.2	$(3.81,-7.36,0.53)$	29358.2	$(3.64,-6.03,0.53)$	28967.6	$(100,100,0.29)$
Bayesian Beliefs						
True supergame (g,l,δ), no free par $(-)$	38912.9	$(-,-,-)$	37205.4	$(-,-,-)$	37901.7	$(-,-,-)$
True stage game g,l, free $(\delta^X,-,-)$	38784.8	$(0.84,-,-)$	37202.4	$(0.98,-,-)$	36329.7	$(1.27,-,-)$
True δ, inequity aversion $(-,\alpha,\beta)$	30829.4	$(-,2.04,1.16)$	32096.1	$(-,2,0.92)$	32411.5	$(-,10.71,0.84)$
Inequity Aversion (δ^X,α,β)	28534.6	$(100,60.49,0.99)$	28609.6	$(58.38,14.71,0.97)$	29106.3	$(2.13,64.75,0.83)$
Cond Cooperation (δ^X,α,β)	33762.8	$(100,1.91,0.15)$	32161.8	$(2.88,2.04,0.5)$	29106	$(11.55,14.75,-0.18)$
Altruism (δ^X,α,β)	29003.9	$(100,-10.62,5.7)$	29047.6	$(17.9,-9.93,5.6)$	29105.4	$(100,-51.77,-0.37)$
Gen Fairness Equilibrium (δ^X,α,β)	32044.4	$(6.93,2.96,0.2)$	30760.7	$(4.79,2.91,0.22)$	29105.3	$(27.06,100,0.28)$

D Information on the experiments re-analyzed

This section provides some background information on the experiments re-analyzed in this paper. Table 12 summarizes and defines the strategies considered by previous studies. Table 13 reviews focus and main results (in terms of identified strategies) of these studies. Table 14 reviews the numbers of subjects and observations, average parameters, and average cooperation rates for all experiments, and Table 15 provides the detailed overview by treatments.

Table 12: Pure strategies considered in behavioral analyses

Strategy	Abbreviation	Description	$(\sigma_{cc},\sigma_{cd},\sigma_{dc},\sigma_{dd})^{\dagger}$	References
Pure Strategies Non-responsive or Memory-1				
Always Defect	AD	Always defects independent of previous outcome	(0,0,0,0)	DF11, DF15, FRD12, FY17, STS13
Always Cooperate	AC	Always cooperates independent of previous outcome	(1,1,1,1) (1,1,0,0)	DF11, DF15, FRD12, B15 FY17, STS13
Grim	G	Only cooperates after cc was last outcome	(1,0,0,0)	AF09, DF11, DF15 FRD12, FY17, STS13
Tit-for-Tat	TFT	Only plays C if opponent did last period	(1,0,1,0)	AF09, DF11, DF15 FRD12, FY17, STS13
Win-stay-Lose-Shift (aka Perfect TFT)	WSLS	Plays same strategy if it was successful, otherwise shifts	(1,0,0,1)	DF11, DF15, FRD12, FY17
False cooperator	C-to-AD	Play c in first round then AD	–	FRD12, FY17
Explorative TFT	D-TFT	Play d in first round then TFT	–	DF15, FRD12, FY17
Alternator	DC-Alt	Play d in first round then alternate c and d	–	FRD12, FY17
Trigger-with-Reversion	GwR	Like Grim but revert to cooperation after cc‡	(1,0,0,0)	STS13
Pure Strategies Memory-2/3				
Trigger 2 periods	T2	Player punishes defection for max. 2 periods, otherwise cooperates	$(1,0,\theta_1^*,0)$	DF11, FY17
Tit-for-2(3)-Tats	TF2T	Defects after 2 defections	$(1,\theta_2,1,\theta_2)$	FRD12, FY17
2-Tits-for-2-Tats	2TF2T	Defects twice after 2 defections	$(1,\theta_3,\theta_3,\theta_3)$	FRD12, FY17
2-Tits-for-1-Tats	2TFT	Defects twice after each defections	$(1,0,\theta_4,0)$	FRD12, FY17
Grim2(3/4)	G2(3)	After 2(3) defections will play D forever	$(1,\theta_5,0,0)$	FRD12, FY17, STS13
Win-stay-Lose-Shift-2	WSLS2	cooperate after (dd,dd),(cc,cc), (dd,cc) otherwise defect	–	FRD12
Explorative TF2(3)T	D-TF2(3)T	Play D in first round then TF2(3)T	–	FRD12, FY17
Explorative Grim2(3)	D-Grim2(3)	Play D in first round then Grim2(3)	–	FRD12, FY17
Behavior Strategies				
Semi-Grim**	SG	Similar to Grim but may cooperate after CD or DC.	$(1,\theta_{SG},\theta_{SG},0)$	B15
Generous TFT	GTFT	Like TFT but cooperate with prob α after CD or DD	$(1,\theta_{GT},1,\theta_{GT})$	FRD12, B15

† σ assigns cooperation probabilities after joint cooperation (cc), unilateral defection by opponent (cd), unilateral defection (dc), and joint defection (dd).
‡ possible if players make mistakes.
* Vector assigning cooperation probabilities $\in \{0,1\}$ depending on the state 2 periods ahead.
** θ_{SG} and θ_{GT} are mixing parameters $\in (0,1)$.

References: **AF09** (Aoyagi and Frechette, 2009), **B15** (Breitmoser, 2015), **DF11** (Dal Bó and Fréchette, 2011), **DF15** (Dal Bó and Fréchette, 2015), **FRD12** (Fudenberg et al., 2012), **FY17** (Fréchette and Yuksel, 2017), **STS13** (Sherstyuk et al., 2013)

Table 13: Overview literature

Reference	Focus	Investigation of Strategies	Strategies found
Aoyagi and Frechette (2009)	Imperfect public monitoring in PD	Mainly avg. coop. rates Mem-1, Mem-2, Threshold	Threshold strat S_0 (same threshold in state 1 & 0)
Blonski et al. (2011)	New δ^* with sucker's payoff	Avg. coop rates	–
Bruttel and Kamecke (2012)	Endgame effects	Elicitation of pure strategies discuss avg. coop. rates	– *
Camera et al. (2012)	Player's strat using finite automata	All possible pure mem-1	large share play unconditional
Dal Bó (2005)	Finitely vs infinitely repeated PD	Avg. cooperation rates	–
Dal Bó and Fréchette (2011)	Players' strategies learning model	selected mem-1 strategies SFEM	AC, AD, TFT
Dal Bó and Fréchette (2015) upd (2017)	Players' strategies	SFEM, elicitation, pure Mem-1, Mem-2 mainly preselected	AD, TFT, Grim
Dreber et al. (2008)	PD extended with punishment option	Agg. cooperation behavior	(AD, Grim, TFT)**
Duffy and Ochs (2009)	Fixed matching of players in PD	Round 1 and avg. coop. rates	–
Fréchette and Yuksel (2017)	De-coupling of expected length of game and discount factor	Avg. coop. rates, SFEM Mem-1, Mem2/3 preselected	Grim, TFT
Fudenberg et al. (2012)	Effect of noise/ uncertainty on leniency	Avg. coop. rate, SFEM, 20 pure Mem-1, Mem-2(3)	AC, AD, Grim, (D)-TFT, 2TFT, Grim2
Kagel and Schley (2013)	Linear payoff transformations	Fist round coop. rates	–
Sherstyuk et al. (2013)	Payment schemes	Avg. cooperation rates, share of correctly predicted actions by selected pure strats	AD, TFT, GwR
Dal Bó and Fréchette (2018)	Determinants of cooperation (meta)	Mainly first round coop	–

* Table 4 column "Strategy" in their study indicating SG in coefficients for cd_{t-1} & cd_{t-2}.
** Reported by Fudenberg et al. (2012).

Table 14: Overview of the data sets used in the analysis

Experiment	Logistics #Subj	Logistics #Dec	Parameters δ	Parameters g	Parameters l	Average cooperation rates $\hat{\sigma}_{\emptyset}$	$\hat{\sigma}_{cc}$		$\hat{\sigma}_{cd}$		$\hat{\sigma}_{dc}$		$\hat{\sigma}_{dd}$
First halves per session													
Aoyagi and Frechette (2009)	38	1650	0.9	0.333	0.111	0.465	0.917	$\gg$	0.45	$\approx$	0.408	$\approx$	0.336
Blonski et al. (2011)	200	3040	0.756	1.345	2.602	0.295	0.89	$\gg$	0.279	$\approx$	0.193	$\gg$	0.034
Bruttel and Kamecke (2012)	36	1920	0.8	1.167	0.833	0.481	0.91	$\gg$	0.286	$\approx$	0.228	$\gg$	0.08
Dal Bó (2005)	102	1320	0.75	0.939	1.061	0.342	0.922	$\gg$	0.212	$<$	0.342	$\gg$	0.089
Dal Bó and Fréchette (2011)	266	17772	0.622	1.062	1.072	0.31	0.951	$\gg$	0.334	$\approx$	0.331	$\gg$	0.063
Dal Bó and Fréchette (2015)	672	22112	0.743	1.579	1.341	0.451	0.94	$\gg$	0.297	$\approx$	0.335	$\gg$	0.057
Dreber et al. (2008)	50	2064	0.75	1.488	1.488	0.488	0.904	$\gg$	0.217	$\approx$	0.213	$\gg$	0.036
Duffy and Ochs (2009)	102	3128	0.9	1	1	0.53	0.904	$\gg$	0.301	$\approx$	0.33	$\gg$	0.111
Fréchette and Yuksel (2017)	50	800	0.75	0.4	0.4	0.737	0.943	$\gg$	0.141	$\approx$	0.266	$\approx$	0.091
Fudenberg et al. (2012)	48	1452	0.875	0.333	0.333	0.756	0.982	$\gg$	0.4	$\approx$	0.427	$\gg$	0.066
Kagel and Schley (2013)	114	7600	0.75	1	0.5	0.573	0.935	$\gg$	0.263	$\approx$	0.295	$\gg$	0.051
Sherstyuk et al. (2013)	56	3052	0.75	1	0.25	0.56	0.945	$\gg$	0.328	$\approx$	0.371	$\gg$	0.117
Pooled	1734	65910	0.728	1.207	1.083	0.389	0.938	$\gg$	0.304	$\approx$	0.322	$\gg$	0.065
Second halves per session													
Aoyagi and Frechette (2009)	38	1400	0.9	0.333	0.111	0.424	0.958	$\gg$	0.398	$\approx$	0.517	$\approx$	0.375
Blonski et al. (2011)	200	5460	0.766	1.282	2.554	0.279	0.923	$\gg$	0.287	$\approx$	0.231	$\gg$	0.02
Bruttel and Kamecke (2012)	36	1632	0.8	1.167	0.833	0.447	0.947	$\gg$	0.221	$\approx$	0.297	$\gg$	0.041
Dal Bó (2005)	102	1650	0.75	0.961	1.039	0.297	0.92	$\gg$	0.242	$<$	0.388	$\gg$	0.064
Dal Bó and Fréchette (2011)	266	19270	0.62	1.122	1.103	0.355	0.979	$\gg$	0.376	$\approx$	0.362	$\gg$	0.041
Dal Bó and Fréchette (2015)	672	29480	0.766	1.666	1.386	0.469	0.976	$\gg$	0.315	$<$	0.402	$\gg$	0.035
Dreber et al. (2008)	50	1838	0.75	1.533	1.533	0.461	0.917	$\gg$	0.128	$\ll$	0.39	$\gg$	0.009
Duffy and Ochs (2009)	102	6018	0.9	1	1	0.684	0.977	$\gg$	0.367	$\approx$	0.391	$\gg$	0.082
Fréchette and Yuksel (2017)	50	1568	0.75	0.4	0.4	0.763	0.97	$\gg$	0.233	$\approx$	0.398	$\gg$	0.069
Fudenberg et al. (2012)	48	1800	0.875	0.333	0.333	0.829	0.971	$\gg$	0.487	$\approx$	0.412	$\gg$	0.083
Kagel and Schley (2013)	114	7172	0.75	1	0.5	0.704	0.966	$\gg$	0.262	$\approx$	0.332	$\gg$	0.025
Sherstyuk et al. (2013)	56	2604	0.75	1	0.25	0.646	0.973	$\gg$	0.482	$\approx$	0.437	$\gg$	0.078
Pooled	1734	79892	0.744	1.271	1.172	0.404	0.971	$\gg$	0.327	$<$	0.376	$\gg$	0.039

Note: The "average cooperation rates" are the relative frequencies estimated directly from the data. The relation signs encode bootstrapped p-values (resampling at the subject level with 10,000 repetitions) where $<,>$ indicate rejection of the Null of equality at $p < .05$ and $\ll,\gg$ indicating $p < .002$. Following Wright (1992), we accommodate for the multiplicity of comparisons within data sets by adjusting p-values using the Holm-Bonferroni method (Holm, 1979). Note that all details here exactly replicate Breitmoser (2015). As a result, if a data set is considered in isolation, the .05-level indicated by "$>,<$" is appropriate. If all 24 treatments are considered simultaneously, the corresponding Bonferroni correction requires to further reduce the threshold to $.002 \approx .05/24$, which corresponds with "$\gg,\ll$".

Table 15: Table 14 by treatments – Overview of the data sets used in the analysis

(a) First halves per session

Treatment	Logistics #Subj	Logistics #Dec	Parameters δ	Parameters g	Parameters l	Average cooperation rates $\hat{\sigma}_\emptyset$	$\hat{\sigma}_{cc}$	$\hat{\sigma}_{cd}$	$\hat{\sigma}_{dc}$	$\hat{\sigma}_{dd}$
Aoyagi and Frechette (2009)										
AF09–34	38	1650	0.9	0.333	0.111	0.729	0.917 $\gg$	0.45 $\approx$	0.408 $\approx$	0.336
Blonski et al. (2011)										
BOS11–9	20	220	0.5	2	2	0.23	-	0.182	0.182	0.031
BOS11–14	20	340	0.75	0.5	3.5	0.16	-	0.188	0.062	0.029
BOS11–15	20	320	0.75	1	8	0.04	-	0.167	0	0.005
BOS11–16	20	400	0.75	0.75	1.25	0.56	0.915 $\gg$	0.206 $\approx$	0.206 $>$	0.073
BOS11–17	20	180	0.75	0.833	0.5	0.42	0.5 $\approx$	0.235 $\approx$	0.471 $\approx$	0.125
BOS11–26	40	760	0.75	2	2	0.285	0.833 $\gg$	0.235 $\approx$	0.196 $\gg$	0.03
BOS11–27	20	240	0.75	1	1	0.28	0.917 $>$	0.316 $\approx$	0.211 $>$	0.056
BOS11–30	20	140	0.875	0.5	3.5	0.275	-	0	0	0.058
BOS11–31	20	440	0.875	2	2	0.437	0.968 $\gg$	0.513 $\approx$	0.154 $>$	0.023
BOS11–All	200	3040	0.756	1.345	2.602	0.295	0.89 $\gg$	0.279 $\approx$	0.193 $\gg$	0.034
Bruttel and Kamecke (2012)										
BK12–28	36	1920	0.8	1.167	0.833	0.481	0.91 $\gg$	0.286 $\approx$	0.228 $\gg$	0.08
Dal Bó (2005)										
D05–18	42	420	0.75	1.167	0.833	0.484	0.806 $\gg$	0.239 $\approx$	0.304 $>$	0.114
D05–19	60	900	0.75	0.833	1.167	0.443	0.958 $\gg$	0.2 $<$	0.36 $\gg$	0.074
D05–All	102	1320	0.75	0.939	1.061	0.342	0.922 $\gg$	0.212 $<$	0.342 $\gg$	0.089
Dal Bó and Fréchette (2011)										
DF11–6	44	2748	0.5	2.571	1.857	0.134	0.792 $\gg$	0.32 $\approx$	0.272 $\gg$	0.036
DF11–7	50	3290	0.5	0.667	0.867	0.18	0.673 $\gg$	0.299 $\approx$	0.258 $\gg$	0.061
DF11–8	46	3092	0.5	0.087	0.565	0.365	0.973 $\gg$	0.421 $>$	0.263 $\gg$	0.081
DF11–22	44	2842	0.75	2.571	1.857	0.248	0.891 $\gg$	0.303 $\approx$	0.355 $\gg$	0.05
DF11–23	38	2656	0.75	0.667	0.867	0.511	0.965 $\gg$	0.39 $\approx$	0.386 $\gg$	0.073
DF11–24	44	3144	0.75	0.087	0.565	0.74	0.961 $\gg$	0.266 $\approx$	0.399 $\gg$	0.11
DF11–All	266	17772	0.622	1.062	1.072	0.31	0.951 $\gg$	0.334 $\approx$	0.331 $\gg$	0.063
Dal Bó and Fréchette (2015)										
DF15–4	50	1438	0.5	2.571	1.857	0.137	0.562 $>$	0.164 $<$	0.327 $\gg$	0.031
DF15–5	140	4094	0.5	0.087	0.565	0.58	0.921 $\gg$	0.254 $\approx$	0.241 $\gg$	0.082
DF15–20	114	4054	0.75	2.571	1.857	0.25	0.912 $\gg$	0.223 $<$	0.336 $\gg$	0.052
DF15–21	164	4740	0.75	0.087	0.565	0.658	0.952 $\gg$	0.388 $\approx$	0.369 $\gg$	0.083
DF15–33	168	6438	0.9	2.571	1.857	0.307	0.928 $\gg$	0.297 $\approx$	0.344 $\gg$	0.054
DF15–35	36	1348	0.95	2.571	1.857	0.5	0.974 $\gg$	0.324 $\approx$	0.432 $\gg$	0.05
DF15–All	672	22112	0.743	1.579	1.341	0.451	0.94 $\gg$	0.297 $\approx$	0.335 $\gg$	0.057
Dreber et al. (2008)										
DRFN08–10	28	1008	0.75	2	2	0.468	0.888 $\gg$	0.188 $\approx$	0.139 $\gg$	0.02
DRFN08–11	22	1056	0.75	1	1	0.507	0.917 $\gg$	0.245 $\approx$	0.283 $\gg$	0.051
DRFN08–All	50	2064	0.75	1.488	1.488	0.488	0.904 $\gg$	0.217 $\approx$	0.213 $\gg$	0.036
Duffy and Ochs (2009)										
DO09–32	102	3128	0.9	1	1	0.53	0.904 $\gg$	0.301 $\approx$	0.33 $\gg$	0.111
Fréchette and Yuksel (2017)										
FY17–25	50	800	0.75	0.4	0.4	0.737	0.943 $\gg$	0.141 $\approx$	0.266 $\approx$	0.091
Fudenberg et al. (2012)										
FRD12–29	48	1452	0.875	0.333	0.333	0.756	0.982 $\gg$	0.4 $\approx$	0.427 $\gg$	0.066
Kagel and Schley (2013)										
KS13–12	114	7600	0.75	1	0.5	0.573	0.935 $\gg$	0.263 $\approx$	0.295 $\gg$	0.051
Sherstyuk et al. (2013)										
STS13–13	56	3052	0.75	1	0.25	0.56	0.945 $\gg$	0.328 $\approx$	0.371 $\gg$	0.117
Pooled	1734	65910	0.728	1.207	1.083	0.389	0.938 $\gg$	0.304 $\approx$	0.322 $\gg$	0.065

(b) Second halves per session

Treatment	Logistics #Subj	Logistics #Dec	Parameters δ	Parameters g	Parameters l	Average cooperation rates $\hat{\sigma}_\emptyset$	$\hat{\sigma}_{cc}$	$\hat{\sigma}_{cd}$	$\hat{\sigma}_{dc}$	$\hat{\sigma}_{dd}$
Aoyagi and Frechette (2009)										
AF09–34	38	1400	0.9	0.333	0.111	0.873	0.958 $\gg$	0.398 $\approx$	0.517 $\approx$	0.375
Blonski et al. (2011)										
BOS11–9	20	300	0.5	2	2	0.233	0.917 $>$	0.062 $\approx$	0.188 $\approx$	0.007
BOS11–14	20	280	0.75	0.5	3.5	0.025	-	0.2	0.4	0.013
BOS11–15	20	640	0.75	1	8	0	-	0	0	0.002
BOS11–16	20	340	0.75	0.75	1.25	0.633	0.846 $\gg$	0.2 $\approx$	0.233 $\gg$	0.024
BOS11–17	20	680	0.75	0.833	0.5	0.417	0.917 $\gg$	0.182 $\approx$	0.255 $\gg$	0.026
BOS11–26	40	1100	0.75	2	2	0.283	0.959 $\gg$	0.241 $\approx$	0.203 $\gg$	0.032
BOS11–27	20	800	0.75	1	1	0.308	0.875 $\gg$	0.447 $\approx$	0.318 $\gg$	0.023
BOS11–30	20	560	0.875	0.5	3.5	0.3	0.8 $\approx$	0.167 $\approx$	0.139 $\approx$	0.02
BOS11–31	20	760	0.875	2	2	0.338	1 $\gg$	0.423 $\approx$	0.173 $>$	0.021
BOS11–All	200	5460	0.766	1.282	2.554	0.279	0.923 $\gg$	0.287 $\approx$	0.231 $\gg$	0.02
Bruttel and Kamecke (2012)										
BK12–28	36	1632	0.8	1.167	0.833	0.447	0.947 $\gg$	0.221 $\approx$	0.297 $\gg$	0.041
Dal Bó (2005)										
D05–18	42	630	0.75	1.167	0.833	0.476	0.86 $\gg$	0.274 $<$	0.476 $\gg$	0.098
D05–19	60	1020	0.75	0.833	1.167	0.533	0.952 $\gg$	0.21 $\approx$	0.296 $\gg$	0.046
D05–All	102	1650	0.75	0.961	1.039	0.297	0.92 $\gg$	0.242 $<$	0.388 $\gg$	0.064
Dal Bó and Fréchette (2011)										
DF11–6	44	2988	0.5	2.571	1.857	0.064	1 $\gg$	0.352 $\approx$	0.477 $\gg$	0.022
DF11–7	50	3614	0.5	0.667	0.867	0.194	0.922 $\gg$	0.377 $\approx$	0.364 $\gg$	0.078
DF11–8	46	3398	0.5	0.087	0.565	0.414	1 $\gg$	0.409 $>$	0.189 $\gg$	0.027
DF11–22	44	3606	0.75	2.571	1.857	0.264	0.96 $\gg$	0.357 $\approx$	0.408 $\gg$	0.024
DF11–23	38	2524	0.75	0.667	0.867	0.708	0.974 $\gg$	0.405 $\approx$	0.5 $\gg$	0.088
DF11–24	44	3140	0.75	0.087	0.565	0.957	0.984 $\gg$	0.302 $\approx$	0.372 $\gg$	0.083
DF11–All	266	19270	0.62	1.122	1.103	0.355	0.979 $\gg$	0.376 $\approx$	0.362 $\gg$	0.041
Dal Bó and Fréchette (2015)										
DF15–4	50	1638	0.5	2.571	1.857	0.101	0.833 $>$	0.067 $<$	0.267 $>$	0.017
DF15–5	140	4656	0.5	0.087	0.565	0.539	0.976 $\gg$	0.27 $\approx$	0.231 $\gg$	0.038
DF15–20	114	4370	0.75	2.571	1.857	0.24	0.948 $\gg$	0.305 $\approx$	0.37 $\gg$	0.03
DF15–21	164	6090	0.75	0.087	0.565	0.775	0.98 $\gg$	0.313 $\approx$	0.313 $\gg$	0.062
DF15–33	168	9718	0.9	2.571	1.857	0.384	0.975 $\gg$	0.314 $\ll$	0.542 $\gg$	0.032
DF15–35	36	3008	0.95	2.571	1.857	0.539	0.981 $\gg$	0.478 $\approx$	0.427 $\gg$	0.039
DF15–All	672	29480	0.766	1.666	1.386	0.469	0.976 $\gg$	0.315 $<$	0.402 $\gg$	0.035
Dreber et al. (2008)										
DRFN08–10	28	980	0.75	2	2	0.269	0.75 $\gg$	0.121 $<$	0.276 $\gg$	0.002
DRFN08–11	22	858	0.75	1	1	0.653	0.942 $\gg$	0.133 $\ll$	0.47 $\gg$	0.028
DRFN08–All	50	1838	0.75	1.533	1.533	0.461	0.917 $\gg$	0.128 $\ll$	0.39 $\gg$	0.009
Duffy and Ochs (2009)										
DO09–32	102	6018	0.9	1	1	0.684	0.977 $\gg$	0.367 $\approx$	0.391 $\gg$	0.082
Fréchette and Yuksel (2017)										
FY17–25	50	1568	0.75	0.4	0.4	0.763	0.97 $\gg$	0.233 $\approx$	0.398 $\gg$	0.069
Fudenberg et al. (2012)										
FRD12–29	48	1800	0.875	0.333	0.333	0.829	0.971 $\gg$	0.487 $\approx$	0.412 $\gg$	0.083
Kagel and Schley (2013)										
KS13–12	114	7172	0.75	1	0.5	0.704	0.966 $\gg$	0.262 $\approx$	0.332 $\gg$	0.025
Sherstyuk et al. (2013)										
STS13–13	56	2604	0.75	1	0.25	0.646	0.973 $\gg$	0.482 $\approx$	0.437 $\gg$	0.078
Pooled	1734	79892	0.744	1.271	1.172	0.404	0.971 $\gg$	0.327 $<$	0.376 $\gg$	0.039

Table 16: Expected and observed realizations in two round 2s per subject after outcome CD in round 1

		Cooperators			Defectors		
		iid	observed	difference	iid	observed	difference
Half 1		(Obs	518)		(Obs	108)	
	Defecting twice	0.557	0.627	-0.07	0.522	0.583	-0.061
	One of each	0.379	0.237	0.142	0.401	0.278	0.123
	Cooperating twice	0.064	0.135	-0.071	0.077	0.139	-0.062
Half 2		(Obs	455)		(Obs	84)	
	Defecting twice	0.557	0.684	-0.116	0.545	0.631	-0.086
	One of each	0.379	0.141	0.23	0.387	0.214	0.173
	Cooperating twice	0.064	0.176	-0.115	0.069	0.155	-0.086

Note: "Cooperators" and "Defectors" are determined by their average cooperation rate in round 1. If above median, they are cooperators. Average cooperation behavior in round 2 if the state is CD of the last two supergames with such an observation by halves and round1-cooperation rates.

Table 17: Overview of cooperation rates in the data

	Cooperators							Defectors						
			Average cooperation rates							Average cooperation rates				
Experiment	#Subj	#Dec	$\hat{\sigma}_{\emptyset}$	$\hat{\sigma}_{cc}$	$\hat{\sigma}_{cd}$	$\hat{\sigma}_{dc}$	$\hat{\sigma}_{dd}$	#Subj	#Dec	$\hat{\sigma}_{\emptyset}$	$\hat{\sigma}_{cc}$	$\hat{\sigma}_{cd}$	$\hat{\sigma}_{dc}$	$\hat{\sigma}_{dd}$
First halves per session														
Aoyagi and Frechette (2009)	35	1509	0.783	0.936	0.45	0.402	0.313	3	141	0.143	0.575	0.444	0.441	0.486
Blonski et al. (2011)	74	1145	0.685	0.896	0.31	0.356	0.056	126	1895	0.066	0.714	0.192	0.123	0.027
Bruttel and Kamecke (2012)	20	1062	0.75	0.926	0.253	0.267	0.113	16	858	0.144	0.806	0.375	0.198	0.055
Dal Bó (2005)	52	675	0.807	0.947	0.21	0.37	0.133	50	645	0.087	0.762	0.22	0.326	0.064
Dal Bó and Fréchette (2011)	108	7382	0.699	0.969	0.337	0.415	0.113	158	10390	0.108	0.807	0.328	0.28	0.045
Dal Bó and Fréchette (2015)	311	10133	0.819	0.954	0.326	0.499	0.084	361	11979	0.124	0.87	0.239	0.239	0.048
Dreber et al. (2008)	31	1272	0.711	0.909	0.189	0.245	0.05	19	792	0.129	0.846	0.326	0.181	0.022
Duffy and Ochs (2009)	63	1886	0.807	0.913	0.302	0.403	0.14	39	1242	0.097	0.866	0.298	0.25	0.087
Fréchette and Yuksel (2017)	41	652	0.886	0.941	0.133	0.394	0.136	9	148	0.056	1	0.25	0.129	0.039
Fudenberg et al. (2012)	39	1185	0.905	0.985	0.418	0.518	0.06	9	267	0.091	0.947	0.316	0.333	0.077
Kagel and Schley (2013)	76	5066	0.814	0.939	0.262	0.419	0.069	38	2534	0.089	0.872	0.268	0.168	0.033
Sherstyuk et al. (2013)	34	1920	0.828	0.968	0.33	0.518	0.119	22	1132	0.152	0.78	0.323	0.266	0.115
Pooled	884	33887	0.778	0.951	0.312	0.43	0.098	850	32023	0.111	0.843	0.283	0.242	0.049
Second halves per session														
Aoyagi and Frechette (2009)	34	1245	0.959	0.968	0.382	0.578	0.328	4	155	0.211	0.75	0.448	0.371	0.469
Blonski et al. (2011)	66	1761	0.75	0.926	0.322	0.398	0.036	134	3699	0.049	0.91	0.189	0.164	0.015
Bruttel and Kamecke (2012)	15	656	0.893	0.954	0.136	0.613	0.031	21	976	0.129	0.922	0.351	0.211	0.044
Dal Bó (2005)	60	974	0.838	0.927	0.24	0.434	0.063	42	676	0.042	0.852	0.25	0.348	0.065
Dal Bó and Fréchette (2011)	111	7984	0.892	0.982	0.358	0.579	0.055	155	11286	0.081	0.948	0.406	0.286	0.038
Dal Bó and Fréchette (2015)	319	14330	0.897	0.978	0.312	0.585	0.067	353	15150	0.089	0.965	0.322	0.315	0.024
Dreber et al. (2008)	22	830	0.847	0.929	0.1	0.479	0.027	28	1008	0.125	0.833	0.195	0.344	0.002
Duffy and Ochs (2009)	69	4206	0.943	0.978	0.376	0.408	0.083	33	1812	0.124	0.968	0.348	0.373	0.081
Fréchette and Yuksel (2017)	42	1322	0.909	0.973	0.227	0.507	0.115	8	246	0	0.8	0.333	0.194	0.014
Fudenberg et al. (2012)	41	1542	0.957	0.969	0.465	0.456	0.106	7	258	0.065	1	0.6	0.325	0.053
Kagel and Schley (2013)	82	5176	0.949	0.968	0.242	0.505	0.035	32	1996	0.067	0.937	0.426	0.194	0.015
Sherstyuk et al. (2013)	37	1674	0.907	0.978	0.489	0.558	0.124	19	930	0.123	0.946	0.456	0.382	0.053
Pooled	898	41700	0.898	0.974	0.318	0.525	0.063	836	38192	0.084	0.954	0.347	0.292	0.03

Note: "Cooperators" and "Defectors" are determined by their average cooperation rate in round 1. If above median, they are cooperators. The "average cooperation rates" are the relative frequencies estimated directly from the data. The relation signs encode bootstrapped p-values (resampling at the subject level with 10,000 repetitions) where $<, >$ indicate rejection of the Null of equality at $p < .05$ and $\ll, \gg$ indicating $p < .002$. Following Wright (1992), we accommodate for the multiplicity of comparisons within data sets by adjusting p-values using the Holm-Bonferroni method (Holm, 1979). Note that all details here exactly replicate Breitmoser (2015). As a result, if a data set is considered in isolation, the .05-level indicated by "$>, <$" is appropriate. If all 24 treatments are considered simultaneously, the corresponding Bonferroni correction requires to further reduce the threshold to $.002 \approx .05/24$, which corresponds with "$\gg, \ll$".

E Robustness checks for Section 4

The tables in this section replicate the tables presented in the Section 3, provide a number of robustness checks and additionally present the results treatment-by-treatment.

- Table 18 compares the "best mixtures" analyzed in the main text to the models allowing for all 1-memory types that correspond with those analyzed in the literature, e.g. Dal Bó and Fréchette (2011). Recall that the 2-memory strategies analyzed in other strings of literature are examined in Section 4. This table clarifies that focussing on the "best mixtures" for each treatment improves the goodness-of-fit of these models substantially (i.e. by at least 100 likelihood points).
- Table 21 compares the best mixtures of pure and generalized pure strategies as discussed in the main text.
- Table 23 is similar to Table 3 in the main text but focussing on the prototypical strategies in their pure form only.
- Table 25 is similar to Table 3 in the main text but focussing on the prototypical strategies in their generalized form only.
- Table 27 is equivalent to Table 3 in the main text.
- Table 29 is equivalent to Table 8 in the main text.
- Table 31 reports on a robustness check for Table 8 by focussing on continuation strategies.
- Table 32 is equivalent to Table 4 in the main text.
- Table 34 shows aggregate state-wise cooperation rates for different lagged histories (cooperation or defection of the opponent in $t-2$) *TFT-Scheme*.
- Table 35 shows aggregate state-wise cooperation rates for different lagged histories (joint cooperation or not in $t-2$) *Grim-Scheme*.
- Table 37 compares different models containing semi-grim to models containing pure strategies assuming no-switching behavior.
- Table 39 compares different models containing semi-grim to models containing pure strategies assuming random-switching behavior.
- Table 41 compares different models containing modifications of semi-grim.
- Table 43 compares different models containing prototypical strategies derived from strategies discussed in previous literature in a No-Switching model.
- Table 45 compares different two parameter versions of semi-grim with models containing prototypical strategies. The memory-2 level follows a *Grim-Scheme* if applicable

- Table 46 compares different two parameter versions of semi-grim with models containing prototypical strategies. The memory-2 level follows a *TFT-Scheme* if applicable
- Table 47 examines all mixtures of Semi-Grim with pure or generalized pure strategies as secondary components.

Table 18: Pure, mixed, or switching strategies? (ICL-BIC of the models, less is better and relation signs point toward better models)

	Best w/o SG						All but SG					
	No Switching		Random Switching		Markov Switching		No Switching		Random Switching		Markov Switching	
Specification												
# Models evaluated	5^{32}		5^{32}		5^{32}		1		1		1	
# Pars estimated (by treatment)	16		16		82		5		5		30	
# Parameters accounted for	3–5		3–5		12–30		5		5		30	
First halves per session												
Aoyagi and Frechette (2009)	843.08	$\approx$	834.4	$\approx$	845.51		886.44	$\approx$	866.95	$\approx$	892.7	
Blonski et al. (2011)	1069.58	$\approx$	1104.85	$\ll$	1221.28		1114.69	$\approx$	1157.02	$\ll$	1615.75	
Bruttel and Kamecke (2012)	845.41	$\approx$	845.05	$>$	785.49		845.41	$\approx$	846.82	$>$	811.17	
Dal Bó (2005)	651.88	$<$	689.58	$>$	652.36		666.1	$<$	702.56	$\approx$	729	
Dal Bó and Fréchette (2011)	7164.32	$\ll$	7557.8	$\gg$	6422.83		7423.23	$<$	7705.11	$\gg$	6913.83	
Dal Bó and Fréchette (2015)	8756.15	$\ll$	9253.62	$\gg$	8275.74		8880.62	$\ll$	9330.5	$\gg$	8571.44	
Dreber et al. (2008)	863.26	$\approx$	864.49	$\gg$	752.16		871.32	$\approx$	880.55	$\gg$	809.71	
Duffy and Ochs (2009)	1396.68	$<$	1467.36	$\gg$	1372.99		1448.71	$<$	1497.48	$>$	1444.51	
Fréchette and Yuksel (2017)	313.03	$\approx$	337.5	$>$	301.74		321.32	$\approx$	337.5	$\approx$	332.73	
Fudenberg et al. (2012)	451.47	$\approx$	435.83	$\approx$	435.86		454.09	$\approx$	437.74	$\approx$	455.21	
Kagel and Schley (2013)	2685.4	$\ll$	3010.1	$\gg$	2439.06		2735.02	$\ll$	3041.29	$\gg$	2581.96	
Sherstyuk et al. (2013)	1346.41	$<$	1481.65	$\gg$	1296.85		1389.33	$<$	1483.17	$\gg$	1333.21	
Pooled	**26525.91**	$\ll$	**28023.06**	$\gg$	**25411.21**		**27218.66**	$\ll$	**28469.06**	$\gg$	**27585.46**	
Second halves per session												
Aoyagi and Frechette (2009)	492.28	$\approx$	484.05	$\approx$	482.82		534.29	$\approx$	514.94	$\approx$	547.48	
Blonski et al. (2011)	1462.41	$\approx$	1513.92	$<$	1604.87		1503.41	$\approx$	1554.93	$\ll$	1973.56	
Bruttel and Kamecke (2012)	561.63	$\approx$	627.74	$\gg$	516.71		588.33	$\approx$	632.75	$>$	584.4	
Dal Bó (2005)	741.2	$<$	790.21	$>$	743.74		751.82	$\ll$	814.54	$\approx$	823.78	
Dal Bó and Fréchette (2011)	5646.38	$\ll$	6634.92	$\gg$	5110.1		6065.93	$\ll$	6783.93	$\gg$	5634.97	
Dal Bó and Fréchette (2015)	8951.57	$\ll$	9835.77	$\gg$	8264.26		9085.4	$\ll$	9876.09	$\gg$	8601.02	
Dreber et al. (2008)	648.55	$\approx$	681.35	$>$	588.62		656.38	$\approx$	702.27	$\approx$	672.66	
Duffy and Ochs (2009)	1925.24	$\approx$	1992.71	$\gg$	1883.22		2010.01	$\approx$	2038.28	$>$	1977.9	
Fréchette and Yuksel (2017)	433.18	$<$	474.93	$>$	427.79		469.85	$\approx$	493.4	$\approx$	474.8	
Fudenberg et al. (2012)	528.36	$\approx$	545.76	$\approx$	529.88		530.3	$\approx$	547.36	$\approx$	549.27	
Kagel and Schley (2013)	1751.81	$\ll$	2365.94	$\gg$	1678.7		1866.19	$\ll$	2375.6	$\gg$	1777.72	
Sherstyuk et al. (2013)	1025.32	$\ll$	1177.96	$\gg$	1008.49		1027.43	$\ll$	1180.11	$\gg$	1025.75	
Pooled	**24301.45**	$\ll$	**27269.48**	$\gg$	**23494.22**		**25271.72**	$\ll$	**27696.6**	$\gg$	**25737.55**	

Note: Relation signs are used as defined above (Table 14). "No Switching", "Random Switching" and "Markov Switching" are as defined in the text, but briefly: "No Switching" assumes that each subject randomly chooses a strategy prior to the first supergame and plays this strategy constantly for the entire half session. "Random Switching" assumes that each subject randomly chooses a strategy prior to each supergame (by i.i.d. draws), and "Markov Switching" allows that these switches follow a Markov process. "All but SG" allows subjects to play either AD, Grim, TFT, AC or WSLS, and "Best w/o SG" picks the best mixture model after eliminating AC or WSLS, or both or none of these.

Table 19: Table 18 by treatments – Pure, mixed, or switching strategies?

(a) First halves per session

	Best w/o SG			All but SG		
	No Switching	Random Switching	Markov Switching	No Switching	Random Switching	Markov Switching
Specification						
# Models evaluated	5^{32}	5^{32}	5^{32}	1	1	1
# Pars estimated (by treatment)	16	16	82	5	5	30
# Parameters accounted for	3–5	3–5	12–30	5	5	30
AF09–34	843.08 $\approx$	834.4 $\approx$	845.51	886.44 $\approx$	866.95 $\approx$	892.7
BOS11–9	83.42 $\approx$	83.96 $\approx$	88.41	85.17 $\approx$	86.21 $\ll$	112.66
BOS11–14	97.73 $\approx$	90 $\approx$	92.94	100.72 $\approx$	93 $\ll$	119.36
BOS11–15	34.3 $\approx$	32.69 $\ll$	43.18	37.29 $\approx$	35.69 $\ll$	69.59
BOS11–16	167.3 $\approx$	169.38 $\approx$	170.57	176.55 $\approx$	180.23 $<$	197.16
BOS11–17	110.57 $\approx$	118.71 $\approx$	121.05	113.57 $\approx$	121.87 $\ll$	147.65
BOS11–26	256.88 $\approx$	262.33 $\approx$	257.54	260.57 $\approx$	270.79 $\approx$	286.48
BOS11–27	102.11 $\approx$	112.76 $\approx$	111.44	103.61 $\approx$	114.26 $\approx$	132.37
BOS11–30	56.81 $\approx$	65.61 $\approx$	64.33	59.81 $\approx$	68.31 $<$	91.33
BOS11–31	125.82 $\approx$	135.1 $\approx$	142.43	127.32 $\approx$	136.59 $<$	158.7
BK12–28	845.41 $\approx$	845.05 $>$	785.49	845.41 $\approx$	846.82 $>$	811.17
D05–18	235.84 $\approx$	234.95 $\approx$	235.63	241.39 $\approx$	243.54 $<$	266.02
D05–19	413.65 $<$	452.05 $\gg$	408.22	421.17 $<$	455.47 $\approx$	441.71
DF11–6	810.5 $<$	925.1 $>$	770.36	880.04 $\approx$	949.19 $\approx$	847.92
DF11–7	1349.47 $\approx$	1364.07 $\gg$	1132.04	1423.93 $\approx$	1388.73 $\gg$	1227.66
DF11–8	1496.25 $\ll$	1712.65 $\gg$	1279.8	1515.51 $<$	1714.28 $\gg$	1316.15
DF11–22	1154.93 $\approx$	1122.94 $\approx$	1066.33	1192.92 $\approx$	1161.02 $\approx$	1141.85
DF11–23	1142.96 $\approx$	1217.02 $\gg$	1020.09	1144.78 $\approx$	1218.9 $\gg$	1066.21
DF11–24	1188.68 $\approx$	1194.48 $\gg$	1046.5	1239.14 $\approx$	1246.06 $>$	1152.48
DF15–4	431.07 $\approx$	467.36 $>$	395.89	460.23 $\approx$	478.56 $\approx$	441.94
DF15–5	1763.19 $\ll$	2211.16 $\gg$	1646.78	1808.3 $\ll$	2212.37 $\gg$	1686.31
DF15–20	1569.49 $\approx$	1543.46 $\gg$	1439.66	1588.62 $\approx$	1571.15 $>$	1501.57
DF15–21	2012.6 $\ll$	2221.98 $\gg$	1943.68	2015.1 $\ll$	2224.97 $\gg$	1960.65
DF15–33	2552.94 $>$	2400.56 $\approx$	2336.73	2573.89 $>$	2423.22 $\approx$	2389.72
DF15–35	403.53 $\approx$	385.77 $\approx$	396.36	405.32 $\approx$	391.08 $\ll$	416.29
DRFN08–10	410.24 $\approx$	390.77 $>$	334.73	413.58 $\approx$	400.1 $>$	362.2
DRFN08–11	450.5 $\approx$	470.91 $\gg$	405.73	454.24 $\approx$	476.95 $\gg$	426.5
DO09–32	1396.68 $<$	1467.36 $\gg$	1372.99	1448.71 $<$	1497.48 $>$	1444.51
FY17–25	313.03 $\approx$	337.5 $>$	301.74	321.32 $\approx$	337.5 $\approx$	332.73
FRD12–29	451.47 $\approx$	435.83 $\approx$	435.86	454.09 $\approx$	437.74 $\approx$	455.21
KS13–12	2685.4 $\ll$	3010.1 $\gg$	2439.06	2735.02 $\ll$	3041.29 $\gg$	2581.96
STS13–13	1346.41 $<$	1481.65 $\gg$	1296.85	1389.33 $<$	1483.17 $\gg$	1333.21
Aoyagi and Frechette (2009)	843.08 $\approx$	834.4 $\approx$	845.51	886.44 $\approx$	866.95 $\approx$	892.7
Blonski et al. (2011)	1069.58 $\approx$	1104.85 $\ll$	1221.28	1114.69 $\approx$	1157.02 $\ll$	1615.75
Bruttel and Kamecke (2012)	845.41 $\approx$	845.05 $>$	785.49	845.41 $\approx$	846.82 $>$	811.17
Dal Bó (2005)	651.88 $<$	689.58 $>$	652.36	666.1 $<$	702.56 $\approx$	729
Dal Bó and Fréchette (2011)	7164.32 $\ll$	7557.8 $\gg$	6422.83	7423.23 $<$	7705.11 $\gg$	6913.83
Dal Bó and Fréchette (2015)	8756.15 $\ll$	9253.62 $\gg$	8275.74	8880.62 $\ll$	9330.5 $\gg$	8571.44
Dreber et al. (2008)	863.26 $\approx$	864.49 $\gg$	752.16	871.32 $\approx$	880.55 $\gg$	809.71
Duffy and Ochs (2009)	1396.68 $<$	1467.36 $\gg$	1372.99	1448.71 $<$	1497.48 $>$	1444.51
Fréchette and Yuksel (2017)	313.03 $\approx$	337.5 $>$	301.74	321.32 $\approx$	337.5 $\approx$	332.73
Fudenberg et al. (2012)	451.47 $\approx$	435.83 $\approx$	435.86	454.09 $\approx$	437.74 $\approx$	455.21
Kagel and Schley (2013)	2685.4 $\ll$	3010.1 $\gg$	2439.06	2735.02 $\ll$	3041.29 $\gg$	2581.96
Sherstyuk et al. (2013)	1346.41 $<$	1481.65 $\gg$	1296.85	1389.33 $<$	1483.17 $\gg$	1333.21
Pooled	**26525.91** $\ll$	**28023.06** $\gg$	**25411.21**	**27218.66** $\ll$	**28469.06** $\gg$	**27585.46**

(b) Second halves per session

	Best w/o SG			All but SG		
	No Switching	Random Switching	Markov Switching	No Switching	Random Switching	Markov Switching
Specification						
# Models evaluated	5^{32}	5^{32}	5^{32}	1	1	1
# Pars estimated (by treatment)	16	16	82	5	5	30
# Parameters accounted for	3–5	3–5	12–30	5	5	30
AF09–34	492.28 $\approx$	484.05 $\approx$	482.82	534.29 $\approx$	514.94 $\approx$	547.48
BOS11–9	84.22 $\approx$	96.42 $\approx$	88.85	87.22 $\approx$	99.64 $\approx$	115.59
BOS11–14	40.82 $\approx$	40.83 $<$	50.24	43.82 $\approx$	43.83 $\ll$	77.2
BOS11–15	15.52 $\approx$	15.52 $\ll$	29.01	18.52 $\approx$	18.52 $\ll$	55.98
BOS11–16	157.48 $\approx$	165.09 $\approx$	157.84	160.48 $\approx$	169.38 $\approx$	183.26
BOS11–17	229.75 $\approx$	225.64 $\approx$	219.73	232.75 $\approx$	230.42 $\approx$	245.01
BOS11–26	366.88 $\approx$	365.76 $\approx$	350.94	369.98 $\approx$	367.47 $\approx$	374.89
BOS11–27	226.92 $\approx$	255.26 $\approx$	243.72	228.41 $\approx$	256.76 $\approx$	258.15
BOS11–30	146.49 $\approx$	137.43 $\approx$	145.96	149.49 $\approx$	143.16 $<$	172.77
BOS11–31	161.17 $\approx$	174.2 $\approx$	173.52	162.67 $\approx$	175.69 $\approx$	190.25
BK12–28	561.63 $\approx$	627.74 $\gg$	516.71	588.33 $\approx$	632.75 $>$	584.4
D05–18	350.59 $\approx$	359.16 $\approx$	351.93	355.62 $\approx$	361.11 $<$	383.77
D05–19	388.49 $<$	428.21 $\gg$	383.3	392.65 $\ll$	449.88 $>$	418.75
DF11–6	633.6 $\approx$	693.84 $\approx$	557.16	751.56 $\approx$	723.62 $\approx$	654.48
DF11–7	1427.15 $<$	1645.34 $\gg$	1268.34	1571.76 $\approx$	1692.09 $\gg$	1428.39
DF11–8	1139.15 $\ll$	1646.78 $\gg$	960.35	1142.1 $\ll$	1648.58 $\gg$	978.71
DF11–22	1196.64 $\approx$	1160.77 $\gg$	1018.52	1198.53 $\approx$	1190.83 $>$	1068.63
DF11–23	723.5 $\ll$	970.63 $\gg$	737.29	842.37 $<$	979.34 $>$	820.55
DF11–24	504.8 $\approx$	496.02 $\approx$	460.73	532.68 $\approx$	522.54 $\approx$	522.66
DF15–4	331.12 $\approx$	402.51 $\approx$	339.15	345.97 $\approx$	407.07 $\approx$	379.36
DF15–5	1666.6 $\ll$	2234.36 $\gg$	1438.87	1686.18 $\ll$	2235.53 $\gg$	1466.72
DF15–20	1572.51 $\approx$	1548.84 $\gg$	1339.13	1572.51 $\approx$	1558.84 $\gg$	1379.3
DF15–21	1664.01 $\ll$	1914.7 $\gg$	1504.63	1754.13 $<$	1914.7 $\gg$	1620.64
DF15–33	2913.27 $\approx$	2919.03 $\gg$	2735.52	2915.83 $\approx$	2936.72 $\gg$	2771.7
DF15–35	779.84 $\approx$	792.29 $\approx$	790.32	781.64 $\approx$	794.07 $\approx$	808.34
DRFN08–10	301.08 $\approx$	289.13 $\approx$	251.55	304.41 $\approx$	303.62 $\approx$	287.94
DRFN08–11	345.37 $\approx$	389.41 $>$	323.06	348.47 $\approx$	395.16 $\approx$	363.7
DO09–32	1925.24 $\approx$	1992.71 $\gg$	1883.22	2010.01 $\approx$	2038.28 $>$	1977.9
FY17–25	433.18 $<$	474.93 $>$	427.79	469.85 $\approx$	493.4 $\approx$	474.8
FRD12–29	528.36 $\approx$	545.76 $\approx$	529.88	530.3 $\approx$	547.36 $\approx$	549.27
KS13–12	1751.81 $\ll$	2365.94 $\gg$	1678.7	1866.19 $\ll$	2375.6 $\gg$	1777.72
STS13–13	1025.32 $\ll$	1177.96 $\gg$	1008.49	1027.43 $\ll$	1180.11 $\gg$	1025.75
Aoyagi and Frechette (2009)	492.28 $\approx$	484.05 $\approx$	482.82	534.29 $\approx$	514.94 $\approx$	547.48
Blonski et al. (2011)	1462.41 $\approx$	1513.92 $<$	1604.87	1503.41 $\approx$	1554.93 $\ll$	1973.56
Bruttel and Kamecke (2012)	561.63 $\approx$	627.74 $\gg$	516.71	588.33 $\approx$	632.75 $>$	584.4
Dal Bó (2005)	741.2 $<$	790.21 $>$	743.74	751.82 $\ll$	814.54 $\approx$	823.78
Dal Bó and Fréchette (2011)	5646.38 $\ll$	6634.92 $\gg$	5110.1	6065.93 $\ll$	6783.93 $\gg$	5634.97
Dal Bó and Fréchette (2015)	8951.57 $\ll$	9835.77 $\gg$	8264.26	9085.4 $\ll$	9876.09 $\gg$	8601.02
Dreber et al. (2008)	648.55 $\approx$	681.35 $>$	588.62	656.38 $\approx$	702.27 $\approx$	672.66
Duffy and Ochs (2009)	1925.24 $\approx$	1992.71 $\gg$	1883.22	2010.01 $\approx$	2038.28 $>$	1977.9
Fréchette and Yuksel (2017)	433.18 $<$	474.93 $>$	427.79	469.85 $\approx$	493.4 $\approx$	474.8
Fudenberg et al. (2012)	528.36 $\approx$	545.76 $\approx$	529.88	530.3 $\approx$	547.36 $\approx$	549.27
Kagel and Schley (2013)	1751.81 $\ll$	2365.94 $\gg$	1678.7	1866.19 $\ll$	2375.6 $\gg$	1777.72
Sherstyuk et al. (2013)	1025.32 $\ll$	1177.96 $\gg$	1008.49	1027.43 $\ll$	1180.11 $\gg$	1025.75
Pooled	**24301.45** $\ll$	**27269.48** $\gg$	**23494.22**	**25271.72** $\ll$	**27696.6** $\gg$	**25737.55**

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 14.

Table 20: Pure, mixed, or switching strategies? Best mixtures of continuation strategies (not including round 1) without Semi-Grim (ICL-BIC of the models, less is better and relation signs point toward better models)

	Best mixture of pure strategies						Best mixture of generalized pure strategies (type II)					
	No Switching		Random Switching		Markov Switching		No Switching		Random Switching		Markov Switching	
Specification												
# Models evaluated	4^{32}		4^{32}		4^{32}		4^{32}		4^{32}		4^{32}	
# Pars estimated (by treatment) (by treatment)	16		16		82		32		32		98	
# Parameters accounted for (by treatment)	3–5		3–5		12–30		6–10		6–10		15–35	
First halves per session												
Aoyagi and Frechette (2009)	744.79	$\approx$	733.65	$\approx$	746.14		645.31	$\approx$	646.53	$\approx$	649.53	
Blonski et al. (2011)	669.18	$\gg$	621.56	$\ll$	843		713.8	$\gg$	670.62	$\ll$	875.74	
Bruttel and Kamecke (2012)	590.68	$\approx$	581	$\approx$	590.89		585.42	$\approx$	570.56	$\approx$	570.87	
Dal Bó (2005)	390.88	$\gg$	363.41	$\ll$	393		407.86	$\gg$	378.95	$<$	404.78	
Dal Bó and Fréchette (2011)	3719.86	$\approx$	3729.53	$\approx$	3670.79		3536.73	$\approx$	3589.36	$\approx$	3524.77	
Dal Bó and Fréchette (2015)	5494.71	$\gg$	5264.13	$\approx$	5303.29		5259.64	$\gg$	5037.82	$\approx$	5057.54	
Dreber et al. (2008)	455.55	$\approx$	461.78	$\approx$	481.64		478.09	$\approx$	466.13	$\approx$	482.86	
Duffy and Ochs (2009)	1069.16	$\approx$	1076.16	$\approx$	1069.38		1047.59	$\approx$	1053.04	$\approx$	1049.79	
Fréchette and Yuksel (2017)	181.98	$\gg$	158.34	$\ll$	176.5		188.5	$\approx$	183.48	$\approx$	175.59	
Fudenberg et al. (2012)	356.73	$>$	331.44	$<$	347.07		319.45	$\approx$	308.6	$\approx$	320.55	
Kagel and Schley (2013)	1776.53	$\approx$	1837.93	$\gg$	1715.12		1761.98	$\approx$	1780.97	$>$	1694.94	
Sherstyuk et al. (2013)	926.9	$\approx$	953.91	$>$	912.67		865.67	$\approx$	907.14	$>$	858.65	
Pooled	**16515.74**	$>$	**16251.31**	$\ll$	**16837.05**		**16077.95**	$>$	**15853.82**	$\ll$	**16335.33**	
Second halves per session												
Aoyagi and Frechette (2009)	448.52	$\approx$	431.35	$\approx$	432.41		363.58	$\approx$	368.23	$\approx$	368.89	
Blonski et al. (2011)	967.16	$\gg$	914.28	$\ll$	1140.5		992.44	$\approx$	993.71	$\ll$	1154.56	
Bruttel and Kamecke (2012)	342.17	$\approx$	361.38	$\approx$	348.88		344.88	$\approx$	358.12	$\approx$	347.08	
Dal Bó (2005)	462.39	$\approx$	445.5	$\ll$	474.71		475.11	$\approx$	456.4	$\approx$	469.98	
Dal Bó and Fréchette (2011)	2957.24	$\approx$	3076.88	$\approx$	2979.53		2737.11	$<$	2875.64	$>$	2721.88	
Dal Bó and Fréchette (2015)	5537.83	$>$	5419.19	$\approx$	5438.75		5164.78	$\approx$	5105.6	$\approx$	5116.42	
Dreber et al. (2008)	287.58	$\approx$	285.34	$<$	303.79		295.06	$\approx$	297.88	$\approx$	303.03	
Duffy and Ochs (2009)	1555.1	$\approx$	1599.27	$\approx$	1561.56		1381.01	$\approx$	1416.71	$\approx$	1392.49	
Fréchette and Yuksel (2017)	333.32	$\approx$	309.06	$\approx$	325.58		309.63	$\approx$	304.7	$\approx$	308.78	
Fudenberg et al. (2012)	443.13	$\approx$	439.28	$\approx$	444.41		373.44	$\approx$	395.32	$\approx$	376.62	
Kagel and Schley (2013)	1191.45	$<$	1301.1	$\gg$	1187.17		1170.12	$\approx$	1224.37	$>$	1143.67	
Sherstyuk et al. (2013)	587.45	$<$	640.1	$>$	597.28		527.09	$\approx$	590.16	$\approx$	567.63	
Pooled	**15249.49**	$\approx$	**15361.1**	$\ll$	**15841.93**		**14387.48**	$<$	**14656.93**	$\approx$	**14961.61**	

Note: Relation signs are used as defined above (Table 14). "No Switching", "Random Switching" and "Markov Switching" are as defined in the text, but briefly: "No Switching" assumes that each subject randomly chooses a strategy prior to the first supergame and plays this strategy constantly for the entire half session. "Random Switching" assumes that each subject randomly chooses a strategy prior to each supergame (by i.i.d. draws), and "Markov Switching" allows that these switches follow a Markov process. "Best mixture of pure strategies" starts with the general mixture model allowing subjects to play AD, Grim, TFT, AC or WSLS and picks the best-fitting model after eliminating AC or WSLS, or both or none of these. The "Best mixture of generalized strategies" additionally allows for randomization based on these proto-typical strategies as defined in the main text.

Table 21: Pure, mixed, or switching strategies? Best mixtures without Semi-Grim, including first round behavior.
(ICL-BIC of the models, less is better and relation signs point toward better models)

	Baseline Model		Best mixture of pure strategies: No Switching		Best mixture of pure strategies: Random Switching		Best mixture of pure strategies: Markov Switching		Best mixture of generalized pure strategies: No Switching		Best mixture of generalized pure strategies: Random Switching		Best mixture of generalized pure strategies: Markov Switching
Specification													
# Models evaluated	1		5^{32}		5^{32}		5^{32}		8^{32}		8^{32}		8^{32}
# Pars estimated (by treatment)	5		16		16		82		64		64		196
# Parameters accounted for	5		3–5		3–5		12–30		6–10		6–10		15–35
First halves per session													
Aoyagi and Frechette (2009)	886.44	$\approx$	843.08	$\approx$	834.4	$\approx$	845.51		756.95	$\approx$	763.11	$\approx$	755.97
Blonski et al. (2011)	1114.69	$\gg$	1069.58	$\approx$	1104.85	$\ll$	1221.28		1134.67	$\approx$	1173.15	$\ll$	1272.13
Bruttel and Kamecke (2012)	845.41	$\approx$	845.41	$\approx$	845.05	$>$	785.49		817.89	$\approx$	835.6	$>$	787.63
Dal Bó (2005)	666.1	$\approx$	651.88	$<$	689.58	$>$	652.36		641.98	$<$	674.57	$\approx$	653.11
Dal Bó and Fréchette (2011)	7423.23	$>$	7164.32	$\ll$	7557.8	$\gg$	6422.83		6921.58	$\ll$	7467.72	$\gg$	6465.99
Dal Bó and Fréchette (2015)	8880.62	$>$	8756.15	$\ll$	9253.62	$\gg$	8275.74		8446	$\ll$	9183.55	$\gg$	8168.2
Dreber et al. (2008)	871.32	$\approx$	863.26	$\approx$	864.49	$\gg$	752.16		787.71	$<$	865.64	$\gg$	763.43
Duffy and Ochs (2009)	1448.71	$\approx$	1396.68	$<$	1467.36	$\gg$	1372.99		1395.4	$<$	1461.01	$>$	1394.31
Fréchette and Yuksel (2017)	321.32	$\approx$	313.03	$\approx$	337.5	$>$	301.74		300.87	$<$	345.74	$>$	298.53
Fudenberg et al. (2012)	454.09	$\approx$	451.47	$\approx$	435.83	$\approx$	435.86		432.32	$\approx$	432.38	$\approx$	425.54
Kagel and Schley (2013)	2735.02	$\approx$	2685.4	$\ll$	3010.1	$\gg$	2439.06		2709.95	$\ll$	2993.4	$\gg$	2539.99
Sherstyuk et al. (2013)	1389.33	$\approx$	1346.41	$<$	1481.65	$\gg$	1296.85		1322.6	$\ll$	1450	$\gg$	1298.37
Pooled	**27218.66**	$\gg$	**26525.91**	$\ll$	**28023.06**	$\gg$	**25411.21**		**25933.42**	$\ll$	**27915.32**	$\gg$	**25504.76**
Second halves per session													
Aoyagi and Frechette (2009)	534.29	$\approx$	492.28	$\approx$	484.05	$\approx$	482.82		416.51	$\approx$	437.8	$\approx$	423.05
Blonski et al. (2011)	1503.41	$\gg$	1462.41	$\approx$	1513.92	$<$	1604.87		1414.39	$\ll$	1553.12	$\approx$	1609.79
Bruttel and Kamecke (2012)	588.33	$\approx$	561.63	$\approx$	627.74	$\gg$	516.71		538.17	$<$	611.91	$\gg$	525.5
Dal Bó (2005)	751.82	$\approx$	741.2	$<$	790.21	$>$	743.74		737.05	$<$	786.21	$>$	741.54
Dal Bó and Fréchette (2011)	6065.93	$>$	5646.38	$\ll$	6634.92	$\gg$	5110.1		5220.17	$\ll$	6378.16	$\gg$	5069.04
Dal Bó and Fréchette (2015)	9085.4	$>$	8951.57	$\ll$	9835.77	$\gg$	8264.26		8205.77	$\ll$	9401.19	$\gg$	7947.33
Dreber et al. (2008)	656.38	$\approx$	648.55	$\approx$	681.35	$>$	588.62		619.9	$\approx$	662.24	$>$	596.78
Duffy and Ochs (2009)	2010.01	$\approx$	1925.24	$\approx$	1992.71	$\gg$	1883.22		1883.52	$\approx$	1914.83	$>$	1850.35
Fréchette and Yuksel (2017)	469.85	$\approx$	433.18	$<$	474.93	$>$	427.79		438.55	$<$	478.2	$\approx$	434.61
Fudenberg et al. (2012)	530.3	$\approx$	528.36	$\approx$	545.76	$\approx$	529.88		514.87	$\approx$	516.12	$\approx$	515.97
Kagel and Schley (2013)	1866.19	$\approx$	1751.81	$\ll$	2365.94	$\gg$	1678.7		1808.21	$\ll$	2336.29	$\gg$	1718.07
Sherstyuk et al. (2013)	1027.43	$\approx$	1025.32	$\ll$	1177.96	$\gg$	1008.49		955.73	$\ll$	1137.49	$\gg$	958.99
Pooled	**25271.72**	$\gg$	**24301.45**	$\ll$	**27269.48**	$\gg$	**23494.22**		**23009.84**	$\ll$	**26479.73**	$\gg$	**23143.38**

Note: Relation signs are used as defined above (Table 14). "No Switching", "Random Switching" and "Markov Switching" are as defined in the text, but briefly: "No Switching" assumes that each subject randomly chooses a strategy prior to the first supergame and plays this strategy constantly for the entire half session. "Random Switching" assumes that each subject randomly chooses a strategy prior to each supergame (by i.i.d. draws), and "Markov Switching" allows that these switches follow a Markov process. "Best mixture of pure strategies" starts with the general mixture model allowing subjects to play AD, Grim, TFT, AC or WSLS and picks the best-fitting model after eliminating AC or WSLS, or both or none of these. The "Best mixture of generalized strategies" additionally allows for randomization in the first round.

Table 22: Table 21 by treatments – Pure, mixed, or switching strategies? Best mixtures without Semi-Grim

(a) First halves per session

	Baseline Model		Best mixture of pure strategies: No Switching		Best mixture of pure strategies: Random Switching		Best mixture of pure strategies: Markov Switching	Best mixture of generalized pure strategies: No Switching		Best mixture of generalized pure strategies: Random Switching		Best mixture of generalized pure strategies: Markov Switching
Specification												
# Models evaluated	1		5^{32}		5^{32}		5^{32}	8^{32}		8^{32}		8^{32}
# Pars estimated (by treatment)	5		16		16		82	64		64		196
# Parameters accounted for	5		3–5		3–5		12–30	6–10		6–10		15–35
AF09–34	886.44	≈	843.08	≈	834.4	≈	845.51	756.95	≈	763.11	≈	755.97
BOS11–9	85.17	≈	83.42	≈	83.96	≈	88.41	89.7	≈	87.81	≈	91.36
BOS11–14	100.72	≈	97.73	≈	90	≈	92.94	102.27	≈	94.44	≈	96.5
BOS11–15	37.29	≈	34.3	≈	32.69	≪	43.18	38.79	≈	37.18	≈	44.74
BOS11–16	176.55	≈	167.3	≈	169.38	≈	170.57	168.92	≈	176.43	≈	174.73
BOS11–17	113.57	≈	110.57	≈	118.71	≈	121.05	115.07	≈	123.19	≈	123.74
BOS11–26	260.57	≈	256.88	≈	262.33	≈	257.54	257.82	≈	269.17	≈	256.37
BOS11–27	103.61	≈	102.11	≈	112.76	≈	111.44	103.44	≈	114.9	≈	110.81
BOS11–30	59.81	>	56.81	≈	65.61	≈	64.33	60.42	≈	68.55	≈	68.16
BOS11–31	127.32	≈	125.82	≈	135.1	≈	142.43	129.65	≈	137.48	≈	145.13
BK12–28	845.41	≈	845.41	≈	845.05	>	785.49	817.89	≈	835.6	>	787.63
D05–18	241.39	≈	235.84	≈	234.95	≈	235.63	241.44	≈	230.66	≈	238.66
D05–19	421.17	≈	413.65	<	452.05	≫	408.22	396.28	≪	439.65	>	403.81
DF11–6	880.04	≈	810.5	<	925.1	>	770.36	823.69	≈	909.31	>	772.55
DF11–7	1423.93	>	1349.47	≈	1364.07	≫	1132.04	1297.64	<	1370.65	≫	1181.36
DF11–8	1515.51	≈	1496.25	≪	1712.65	≫	1279.8	1422.73	≪	1668.83	≫	1284.25
DF11–22	1192.92	≈	1154.93	≈	1122.94	≈	1066.33	1080.23	≈	1110.68	≈	1056.77
DF11–23	1144.78	≈	1142.96	≈	1217.02	≫	1020.09	1082.68	<	1185.69	≫	1027.3
DF11–24	1239.14	≈	1188.68	≈	1194.48	≫	1046.5	1171.57	≈	1179.6	≫	1022.62
DF15–4	460.23	>	431.07	≈	467.36	>	395.89	439.54	≈	474.37	≈	412.27
DF15–5	1808.3	≈	1763.19	≪	2211.16	≫	1646.78	1762.23	≪	2211.09	≫	1638.92
DF15–20	1588.62	≈	1569.49	≈	1543.46	≫	1439.66	1463.03	<	1547.14	≫	1433.87
DF15–21	2015.1	≈	2012.6	≪	2221.98	≫	1943.68	1974.94	≪	2184.97	≫	1902.95
DF15–33	2573.89	≈	2552.94	>	2400.56	≈	2336.73	2379.17	≈	2350.87	≈	2296.41
DF15–35	405.32	≈	403.53	≈	385.77	≈	396.36	384.6	≈	372.62	≈	382.07
DRFN08–10	413.58	≈	410.24	≈	390.77	>	334.73	374.3	≈	391.56	>	339.73
DRFN08–11	454.24	≈	450.5	≈	470.91	≫	405.73	408.63	<	468.48	≫	413.19
DO09–32	1448.71	≈	1396.68	<	1467.36	≫	1372.99	1395.4	<	1461.01	>	1394.31
FY17–25	321.32	≈	313.03	≈	337.5	>	301.74	300.87	<	345.74	>	298.53
FRD12–29	454.09	≈	451.47	≈	435.83	≈	435.86	432.32	≈	432.38	≈	425.54
KS13–12	2735.02	≈	2685.4	≪	3010.1	≫	2439.06	2709.95	≪	2993.4	≫	2539.99
STS13–13	1389.33	≈	1346.41	<	1481.65	≫	1296.85	1322.6	≪	1450	≫	1298.37
Aoyagi and Frechette (2009)	886.44	≈	843.08	≈	834.4	≈	845.51	756.95	≈	763.11	≈	755.97
Blonski et al. (2011)	1114.69	≫	1069.58	≈	1104.85	≪	1221.28	1134.67	≈	1173.15	≪	1272.13
Bruttel and Kamecke (2012)	845.41	≈	845.41	≈	845.05	>	785.49	817.89	≈	835.6	>	787.63
Dal Bó (2005)	666.1	≈	651.88	<	689.58	>	652.36	641.98	<	674.57	≈	653.11
Dal Bó and Fréchette (2011)	7423.23	>	7164.32	≪	7557.8	≫	6422.83	6921.58	≪	7467.72	≫	6465.99
Dal Bó and Fréchette (2015)	8880.62	>	8756.15	≪	9253.62	≫	8275.74	8446	≪	9183.55	≫	8168.2
Dreber et al. (2008)	871.32	≈	863.26	≈	864.49	≫	752.16	787.71	<	865.64	≫	763.43
Duffy and Ochs (2009)	1448.71	≈	1396.68	<	1467.36	≫	1372.99	1395.4	<	1461.01	>	1394.31
Fréchette and Yuksel (2017)	321.32	≈	313.03	≈	337.5	>	301.74	300.87	<	345.74	>	298.53
Fudenberg et al. (2012)	454.09	≈	451.47	≈	435.83	≈	435.86	432.32	≈	432.38	≈	425.54
Kagel and Schley (2013)	2735.02	≈	2685.4	≪	3010.1	≫	2439.06	2709.95	≪	2993.4	≫	2539.99
Sherstyuk et al. (2013)	1389.33	≈	1346.41	<	1481.65	≫	1296.85	1322.6	≪	1450	≫	1298.37
Pooled	**27218.66**	≫	**26525.91**	≪	**28023.06**	≫	**25411.21**	**25933.42**	≪	**27915.32**	≫	**25504.76**

(b) Second halves per session

	Baseline Model		Best mixture of pure strategies: No Switching		Best mixture of pure strategies: Random Switching		Best mixture of pure strategies: Markov Switching	Best mixture of generalized pure strategies: No Switching		Best mixture of generalized pure strategies: Random Switching		Best mixture of generalized pure strategies: Markov Switching
Specification												
# Models evaluated	1		5^{32}		5^{32}		5^{32}	8^{32}		8^{32}		8^{32}
# Pars estimated (by treatment)	5		16		16		82	64		64		196
# Parameters accounted for	5		3–5		3–5		12–30	6–10		6–10		15–35
AF09–34	534.29	≈	492.28	≈	484.05	≈	482.82	416.51	≈	437.8	≈	423.05
BOS11–9	87.22	≈	84.22	≈	96.42	≈	88.85	78.84	<	103.47	≈	83.98
BOS11–14	43.82	≈	40.82	≈	40.83	<	50.24	45.31	≈	42.43	≈	48.97
BOS11–15	18.52	≈	15.52	≈	15.52	≪	29.01	20.01	≈	20.01	≪	33.5
BOS11–16	160.48	≈	157.48	≈	165.09	≈	157.84	148.98	≈	168.12	≈	158.84
BOS11–17	232.75	≈	229.75	≈	225.64	≈	219.73	211.59	≈	225.1	≈	216.6
BOS11–26	369.98	≈	366.88	≈	365.76	≈	350.94	327.16	≈	352.05	≈	338.09
BOS11–27	228.41	≈	226.92	≈	255.26	≈	243.72	224.85	≈	254.56	≈	233.57
BOS11–30	149.49	>	146.49	≈	137.43	≈	145.96	139.46	≈	139.47	≈	146.9
BOS11–31	162.67	≈	161.17	≈	174.2	≈	173.52	151.87	<	179.31	≈	171.15
BK12–28	588.33	≈	561.63	≈	627.74	≫	516.71	538.17	<	611.91	≫	525.5
D05–18	355.62	≈	350.59	≈	359.16	≈	351.93	340.33	≈	355.81	≈	346.45
D05–19	392.65	≈	388.49	<	428.21	≫	383.3	392.47	<	426.15	>	384.46
DF11–6	751.56	≈	633.6	≈	693.84	≈	557.16	579.84	≈	628.84	≈	565.43
DF11–7	1571.76	>	1427.15	<	1645.34	≫	1268.34	1359.89	≪	1582.11	≫	1299.86
DF11–8	1142.1	≈	1139.15	≪	1646.78	≫	960.35	1028.93	≪	1600.86	≫	904.89
DF11–22	1198.53	≈	1196.64	≈	1160.77	≫	1018.52	1012.26	<	1102.07	≫	973.62
DF11–23	842.37	≈	723.5	≪	970.63	≫	737.29	743.89	<	943.35	≫	739.29
DF11–24	532.68	≈	504.8	≈	496.02	≈	460.73	450.61	≈	477.85	≈	455.62
DF15–4	345.97	≈	331.12	≈	402.51	≈	339.15	301.69	<	385.5	≈	307.03
DF15–5	1686.18	≈	1666.6	≪	2234.36	≫	1438.87	1581.28	≪	2217.39	≫	1435.63
DF15–20	1572.51	≈	1572.51	≈	1548.84	≫	1339.13	1273.14	≪	1441.16	≫	1270.1
DF15–21	1754.13	≈	1664.01	≪	1914.7	≫	1504.63	1688.09	<	1878.92	≫	1544.66
DF15–33	2915.83	≈	2913.27	≈	2919.03	≫	2735.52	2582.61	<	2690.87	≫	2541.5
DF15–35	781.64	≈	779.84	≈	792.29	≈	790.32	733.28	≈	742.93	≈	723.78
DRFN08–10	304.41	≈	301.08	≈	289.13	≈	251.55	276.61	≈	285.26	≈	243.71
DRFN08–11	348.47	≈	345.37	≈	389.41	>	323.06	339.09	≈	371.38	≈	339.95
DO09–32	2010.01	≈	1925.24	≈	1992.71	≫	1883.22	1883.52	≈	1914.83	>	1850.35
FY17–25	469.85	≈	433.18	<	474.93	>	427.79	438.55	<	478.2	≈	434.61
FRD12–29	530.3	≈	528.36	≈	545.76	≈	529.88	514.87	≈	516.12	≈	515.97
KS13–12	1866.19	≈	1751.81	≪	2365.94	≫	1678.7	1808.21	≪	2336.29	≫	1718.07
STS13–13	1027.43	≈	1025.32	≪	1177.96	≫	1008.49	955.73	≪	1137.49	≫	958.99
Aoyagi and Frechette (2009)	534.29	≈	492.28	≈	484.05	≈	482.82	416.51	≈	437.8	≈	423.05
Blonski et al. (2011)	1503.41	≫	1462.41	≈	1513.92	<	1604.87	1414.39	≪	1553.12	≈	1609.79
Bruttel and Kamecke (2012)	588.33	≈	561.63	≈	627.74	≫	516.71	538.17	<	611.91	≫	525.5
Dal Bó (2005)	751.82	≈	741.2	<	790.21	>	743.74	737.05	<	786.21	>	741.54
Dal Bó and Fréchette (2011)	6065.93	>	5646.38	≪	6634.92	≫	5110.1	5220.17	≪	6378.16	≫	5069.04
Dal Bó and Fréchette (2015)	9085.4	>	8951.57	≪	9835.77	≫	8264.26	8205.77	≪	9401.19	≫	7947.33
Dreber et al. (2008)	656.38	≈	648.55	≈	681.35	>	588.62	619.9	≈	662.24	>	596.78
Duffy and Ochs (2009)	2010.01	≈	1925.24	≈	1992.71	≫	1883.22	1883.52	≈	1914.83	>	1850.35
Fréchette and Yuksel (2017)	469.85	≈	433.18	<	474.93	>	427.79	438.55	<	478.2	≈	434.61
Fudenberg et al. (2012)	530.3	≈	528.36	≈	545.76	≈	529.88	514.87	≈	516.12	≈	515.97
Kagel and Schley (2013)	1866.19	≈	1751.81	≪	2365.94	≫	1678.7	1808.21	≪	2336.29	≫	1718.07
Sherstyuk et al. (2013)	1027.43	≈	1025.32	≪	1177.96	≫	1008.49	955.73	≪	1137.49	≫	958.99
Pooled	**25271.72**	≫	**24301.45**	≪	**27269.48**	≫	**23494.22**	**23009.84**	≪	**26479.73**	≫	**23143.38**

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 14.

Table 23: Best mixtures of pure strategies in relation to a Semi-Grim behavior strategy
(ICL-BIC of the models, less is better and relation signs point toward better models)

	Best mixture of pure strategies										
	No Switching		Random Switching		Markov Switching		Best Switching		Semi-Grim		AD + SG
Specification											
# Models evaluated	5^{32}		5^{32}		5^{32}				1		1
# Pars estimated (by treatment)	16		16		82				3		5
# Parameters accounted for	3–5		3–5		12–35				3		5
First halves per session											
Aoyagi and Frechette (2009)	843.08	$\approx$	834.4	$\approx$	845.51		845.51	$\gg$	781.86	$\approx$	792.51
Blonski et al. (2011)	1069.58	$\approx$	1104.85	$\ll$	1221.28		1221.28	$\gg$	1069.28	$\approx$	1104.6
Bruttel and Kamecke (2012)	845.41	$\approx$	845.05	$>$	785.49		785.49	$\approx$	800.12	$\approx$	771.14
Dal Bó (2005)	651.88	$<$	689.58	$>$	652.36		652.36	$\approx$	629.17	$\approx$	618.39
Dal Bó and Fréchette (2011)	7164.32	$\ll$	7557.8	$\gg$	6422.83		6422.83	$<$	6597.93	$>$	6352.59
Dal Bó and Fréchette (2015)	8756.15	$\ll$	9253.62	$\gg$	8275.74		8275.74	$>$	8017.59	$\gg$	7830.12
Dreber et al. (2008)	863.26	$\approx$	864.49	$\gg$	752.16		752.16	$\approx$	782.37	$\approx$	764.44
Duffy and Ochs (2009)	1396.68	$<$	1467.36	$\gg$	1372.99		1372.99	$\approx$	1372.97	$\approx$	1361.15
Fréchette and Yuksel (2017)	313.03	$\approx$	337.5	$>$	301.74		301.74	$\approx$	299.62	$\approx$	289.54
Fudenberg et al. (2012)	451.47	$\approx$	435.83	$\approx$	435.86		435.86	$>$	381.01	$\approx$	377.96
Kagel and Schley (2013)	2685.4	$\ll$	3010.1	$\gg$	2439.06		2439.06	$\approx$	2561.76	$\gg$	2450.24
Sherstyuk et al. (2013)	1346.41	$<$	1481.65	$\gg$	1296.85		1296.85	$\approx$	1303.8	$\approx$	1234.52
Pooled	**26525.91**	$\ll$	**28023.06**	$\gg$	**25411.21**		**25411.21**	$\gg$	**24779.85**	$\gg$	**24202.51**
Second halves per session											
Aoyagi and Frechette (2009)	492.28	$\approx$	484.05	$\approx$	482.82		492.28	$\gg$	423.68	$\approx$	421.21
Blonski et al. (2011)	1462.41	$\approx$	1513.92	$<$	1604.87		1462.41	$\gg$	1346.79	$\approx$	1370.16
Bruttel and Kamecke (2012)	561.63	$\approx$	627.74	$\gg$	516.71		561.63	$\approx$	536.77	$\gg$	480.47
Dal Bó (2005)	741.2	$<$	790.21	$>$	743.74		741.2	$>$	699.05	$\approx$	677.24
Dal Bó and Fréchette (2011)	5646.38	$\ll$	6634.92	$\gg$	5110.1		5646.38	$\gg$	5128.69	$\gg$	4565.93
Dal Bó and Fréchette (2015)	8951.57	$\ll$	9835.77	$\gg$	8264.26		8951.57	$\gg$	7825.98	$\gg$	7306.25
Dreber et al. (2008)	648.55	$\approx$	681.35	$>$	588.62		648.55	$>$	589.84	$>$	544.66
Duffy and Ochs (2009)	1925.24	$\approx$	1992.71	$\gg$	1883.22		1925.24	$>$	1761.6	$\gg$	1656.55
Fréchette and Yuksel (2017)	433.18	$<$	474.93	$>$	427.79		433.18	$\approx$	423.34	$\approx$	422.61
Fudenberg et al. (2012)	528.36	$\approx$	545.76	$\approx$	529.88		528.36	$\gg$	452.6	$\approx$	433.74
Kagel and Schley (2013)	1751.81	$\ll$	2365.94	$\gg$	1678.7		1751.81	$\approx$	1775.62	$\gg$	1572.95
Sherstyuk et al. (2013)	1025.32	$\ll$	1177.96	$\gg$	1008.49		1025.32	$\approx$	951.34	$\gg$	834.74
Pooled	**24301.45**	$\ll$	**27269.48**	$\gg$	**23494.22**		**24301.45**	$\gg$	**22097.67**	$\gg$	**20541.83**

Note: This table extends Table 18 by picking the best switching model per half-session, after picking the best-fitting mixture involving the pure forms of AD, Grim, TFT, AC and WSLS (as above) for each treatment independently, and examining its goodness-of-fit in relation to Semi-Grim and mixtures involving Semi-Grim. The model "AD+SG2" has the same number of degrees of freedom as the Semi-Grim model. In contrast to "Semi-Grim", SG2 has only two decrees of freedom including the noise term.

Table 24: Table 23 by treatments – Best mixtures of pure strategies in relation to a Semi-Grim behavior strategy

(a) First halves per session

	Best mixture of pure strategies											
	No Switching		Random Switching		Markov Switching		Best Switching		Semi-Grim		AD + SG	
Specification												
# Models evaluated	5^{32}		5^{32}		5^{32}				1		1	
# Pars estimated (by treatment)	16		16		82				3		5	
# Parameters accounted for	3–5		3–5		12–35				3		5	
AF09–34	843.08	≈	834.4	≈	845.51	845.51	≫	781.86	≈	792.51		
BOS11–9	83.42	≈	83.96	≈	88.41	88.41	≈	86.56	≈	88.35		
BOS11–14	97.73	≈	90	≈	92.94	92.94	≈	93.88	≈	98.01		
BOS11–15	34.3	≈	32.69	≪	43.18	43.18	>	37.73	<	43.07		
BOS11–16	167.3	≈	169.38	≈	170.57	170.57	≈	167.42	≈	157.13		
BOS11–17	110.57	≈	118.71	≈	121.05	121.05	>	115.02	≈	119.79		
BOS11–26	256.88	≈	262.33	≈	257.54	257.54	≈	244.5	≈	246.46		
BOS11–27	102.11	≈	112.76	≈	111.44	111.44	>	92.83	≈	92.07		
BOS11–30	56.81	≈	65.61	≈	64.33	64.33	>	55.74	<	61.12		
BOS11–31	125.82	≈	135.1	≈	142.43	142.43	≈	125.52	≈	128.49		
BK12–28	845.41	≈	845.05	>	785.49	785.49	≈	800.12	≈	771.14		
D05–18	235.84	≈	234.95	≈	235.63	235.63	≈	230.57	≈	238.35		
D05–19	413.65	<	452.05	≫	408.22	408.22	≈	395.06	≈	375.07		
DF11–6	810.5	<	925.1	>	770.36	770.36	≈	794.44	≈	748.9		
DF11–7	1349.47	≈	1364.07	≫	1132.04	1132.04	<	1256.83	≈	1229.61		
DF11–8	1496.25	≪	1712.65	≫	1279.8	1279.8	<	1389.06	≈	1286.68		
DF11–22	1154.93	≈	1122.94	≈	1066.33	1066.33	>	965.96	≈	961.58		
DF11–23	1142.96	≈	1217.02	≫	1020.09	1020.09	≈	1019.57	≈	972.17		
DF11–24	1188.68	≈	1194.48	≫	1046.5	1046.5	<	1145.15	≈	1115.95		
DF15–4	431.07	≈	467.36	>	395.89	395.89	≈	436.57	≈	422.43		
DF15–5	1763.19	≪	2211.16	≫	1646.78	1646.78	<	1738.77	≫	1649.1		
DF15–20	1569.49	≈	1543.46	≫	1439.66	1439.66	≈	1405.3	>	1366.09		
DF15–21	2012.6	≪	2221.98	≫	1943.68	1943.68	>	1872.47	>	1827.37		
DF15–33	2552.94	>	2400.56	≈	2336.73	2336.73	>	2186.26	≈	2178.12		
DF15–35	403.53	≈	385.77	≈	396.36	396.36	>	349.06	≈	346.18		
DRFN08–10	410.24	≈	390.77	>	334.73	334.73	<	367.86	≈	359.04		
DRFN08–11	450.5	≈	470.91	≫	405.73	405.73	≈	411.01	≈	400.49		
DO09–32	1396.68	<	1467.36	≫	1372.99	1372.99	≈	1372.97	≈	1361.15		
FY17–25	313.03	≈	337.5	>	301.74	301.74	≈	299.62	≈	289.54		
FRD12–29	451.47	≈	435.83	≈	435.86	435.86	>	381.01	≈	377.96		
KS13–12	2685.4	≪	3010.1	≫	2439.06	2439.06	≈	2561.76	≫	2450.24		
STS13–13	1346.41	<	1481.65	≫	1296.85	1296.85	≈	1303.8	≈	1234.52		
Aoyagi and Frechette (2009)	843.08	≈	834.4	≈	845.51	845.51	≫	781.86	≈	792.51		
Blonski et al. (2011)	1069.58	≈	1104.85	≪	1221.28	1221.28	≫	1069.28	≈	1104.6		
Bruttel and Kamecke (2012)	845.41	≈	845.05	>	785.49	785.49	≈	800.12	≈	771.14		
Dal Bó (2005)	651.88	<	689.58	>	652.36	652.36	≈	629.17	≈	618.39		
Dal Bó and Fréchette (2011)	7164.32	≪	7557.8	≫	6422.83	6422.83	<	6597.93	>	6352.59		
Dal Bó and Fréchette (2015)	8756.15	≪	9253.62	≫	8275.74	8275.74	>	8017.59	≫	7830.12		
Dreber et al. (2008)	863.26	≈	864.49	≫	752.16	752.16	≈	782.37	≈	764.44		
Duffy and Ochs (2009)	1396.68	<	1467.36	≫	1372.99	1372.99	≈	1372.97	≈	1361.15		
Fréchette and Yuksel (2017)	313.03	≈	337.5	>	301.74	301.74	≈	299.62	≈	289.54		
Fudenberg et al. (2012)	451.47	≈	435.83	≈	435.86	435.86	>	381.01	≈	377.96		
Kagel and Schley (2013)	2685.4	≪	3010.1	≫	2439.06	2439.06	≈	2561.76	≫	2450.24		
Sherstyuk et al. (2013)	1346.41	<	1481.65	≫	1296.85	1296.85	≈	1303.8	≈	1234.52		
Pooled	**26525.91**	≪	**28023.06**	≫	**25411.21**	**25411.21**	≫	**24779.85**	≫	**24202.51**		

(b) Second halves per session

	Best mixture of pure strategies											
	No Switching		Random Switching		Markov Switching		Best Switching		Semi-Grim		AD + SG	
Specification												
# Models evaluated	5^{32}		5^{32}		5^{32}				1		1	
# Pars estimated (by treatment)	16		16		82				3		5	
# Parameters accounted for	3–5		3–5		12–35				3		5	
AF09–34	492.28	≈	484.05	≈	482.82	492.28	≫	423.68	≈	421.21		
BOS11–9	84.22	≈	96.42	≈	88.85	84.22	≈	75.1	≈	80.12		
BOS11–14	40.82	≈	40.83	<	50.24	40.82	≈	35.58	≈	35.86		
BOS11–15	15.52	≈	15.52	≪	29.01	15.52	≈	19.23	<	24.71		
BOS11–16	157.48	≈	165.09	≈	157.84	157.48	≈	150.95	≈	138.89		
BOS11–17	229.75	≈	225.64	≈	219.73	229.75	>	196.25	≈	201.03		
BOS11–26	366.88	≈	365.76	≈	350.94	366.88	>	299.85	≈	309.63		
BOS11–27	226.92	≈	255.26	≈	243.72	226.92	≈	235.88	≈	223.91		
BOS11–30	146.49	≈	137.43	≈	145.96	146.49	≈	129.86	≈	132.45		
BOS11–31	161.17	≈	174.2	≈	173.52	161.17	≈	154.02	≈	153.45		
BK12–28	561.63	≈	627.74	≫	516.71	561.63	≈	536.77	≫	480.47		
D05–18	350.59	≈	359.16	≈	351.93	350.59	≈	334.18	>	312.74		
D05–19	388.49	<	428.21	≫	383.3	388.49	>	361.33	≈	359.54		
DF11–6	633.6	≈	693.84	≈	557.16	633.6	>	526.15	≈	489.74		
DF11–7	1427.15	<	1645.34	≫	1268.34	1427.15	≈	1316.79	≈	1250.02		
DF11–8	1139.15	≪	1646.78	≫	960.35	1139.15	≈	1078.24	>	871.84		
DF11–22	1196.64	≈	1160.77	≫	1018.52	1196.64	≫	930.3	>	858.03		
DF11–23	723.5	≪	970.63	≫	737.29	723.5	≈	767.04	>	608.87		
DF11–24	504.8	≈	496.02	≈	460.73	504.8	≈	483.25	≈	449.72		
DF15–4	331.12	≈	402.51	≈	339.15	331.12	≈	320.02	≈	299.49		
DF15–5	1666.6	≪	2234.36	≫	1438.87	1666.6	≈	1606.96	≫	1407.26		
DF15–20	1572.51	≈	1548.84	≫	1339.13	1572.51	≫	1232.33	≫	1145.96		
DF15–21	1664.01	≪	1914.7	≫	1504.63	1664.01	≈	1591.27	≫	1453.74		
DF15–33	2913.27	≈	2919.03	≫	2735.52	2913.27	≫	2405.23	≈	2331.38		
DF15–35	779.84	≈	792.29	≈	790.32	779.84	≫	641.02	≈	627.6		
DRFN08–10	301.08	≈	289.13	≈	251.55	301.08	>	244.91	≈	234.76		
DRFN08–11	345.37	≈	389.41	>	323.06	345.37	≈	341.42	>	305		
DO09–32	1925.24	≈	1992.71	≫	1883.22	1925.24	>	1761.6	≫	1656.55		
FY17–25	433.18	<	474.93	>	427.79	433.18	≈	423.34	≈	422.61		
FRD12–29	528.36	≈	545.76	≈	529.88	528.36	≫	452.6	≈	433.74		
KS13–12	1751.81	≪	2365.94	≫	1678.7	1751.81	≈	1775.62	≫	1572.95		
STS13–13	1025.32	≪	1177.96	≫	1008.49	1025.32	≈	951.34	≫	834.74		
Aoyagi and Frechette (2009)	492.28	≈	484.05	≈	482.82	492.28	≫	423.68	≈	421.21		
Blonski et al. (2011)	1462.41	≈	1513.92	<	1604.87	1462.41	≫	1346.79	≈	1370.16		
Bruttel and Kamecke (2012)	561.63	≈	627.74	≫	516.71	561.63	≈	536.77	≫	480.47		
Dal Bó (2005)	741.2	<	790.21	>	743.74	741.2	>	699.05	≈	677.24		
Dal Bó and Fréchette (2011)	5646.38	≪	6634.92	≫	5110.1	5646.38	≫	5128.69	≫	4565.93		
Dal Bó and Fréchette (2015)	8951.57	≪	9835.77	≫	8264.26	8951.57	≫	7825.98	≫	7306.25		
Dreber et al. (2008)	648.55	≈	681.35	>	588.62	648.55	>	589.84	>	544.66		
Duffy and Ochs (2009)	1925.24	≈	1992.71	≫	1883.22	1925.24	>	1761.6	≫	1656.55		
Fréchette and Yuksel (2017)	433.18	<	474.93	>	427.79	433.18	≈	423.34	≈	422.61		
Fudenberg et al. (2012)	528.36	≈	545.76	≈	529.88	528.36	≫	452.6	≈	433.74		
Kagel and Schley (2013)	1751.81	≪	2365.94	≫	1678.7	1751.81	≈	1775.62	≫	1572.95		
Sherstyuk et al. (2013)	1025.32	≪	1177.96	≫	1008.49	1025.32	≈	951.34	≫	834.74		
Pooled	**24301.45**	≪	**27269.48**	≫	**23494.22**	**24301.45**	≫	**22097.67**	≫	**20541.83**		

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 14.

Table 25: Best mixtures of generalized strategies in relation to a Semi-Grim strategy
(ICL-BIC of the models, less is better and relation signs point toward better models)

	Best mixture of generalized strategies										Semi-Grim		AD + SG
	No Switching		Random Switching		Markov Switching		Best Switching						
Specification													
# Models evaluated	8^{32}		8^{32}		8^{32}						1		1
# Pars estimated (by treatment)	64		64		196						3		5
# Parameters accounted for	6–10		6–10		15-35						3		5
First halves per session													
Aoyagi and Frechette (2009)	756.95	$\approx$	763.11	$\approx$	755.97		755.97	$\approx$			781.86	$\approx$	792.51
Blonski et al. (2011)	1134.67	$\approx$	1173.15	$\ll$	1272.13		1272.13	$\gg$			1069.28	$\approx$	1104.6
Bruttel and Kamecke (2012)	817.89	$\approx$	835.6	$>$	787.63		787.63	$\approx$			800.12	$\approx$	771.14
Dal Bó (2005)	641.98	$<$	674.57	$\approx$	653.11		653.11	$\approx$			629.17	$\approx$	618.39
Dal Bó and Fréchette (2011)	6921.58	$\ll$	7467.72	$\gg$	6465.99		6465.99	$\approx$			6597.93	$>$	6352.59
Dal Bó and Fréchette (2015)	8446	$\ll$	9183.55	$\gg$	8168.2		8168.2	$>$			8017.59	$\gg$	7830.12
Dreber et al. (2008)	787.71	$<$	865.64	$\gg$	763.43		763.43	$\approx$			782.37	$\approx$	764.44
Duffy and Ochs (2009)	1395.4	$<$	1461.01	$>$	1394.31		1394.31	$\approx$			1372.97	$\approx$	1361.15
Fréchette and Yuksel (2017)	300.87	$<$	345.74	$>$	298.53		298.53	$\approx$			299.62	$\approx$	289.54
Fudenberg et al. (2012)	432.32	$\approx$	432.38	$\approx$	425.54		425.54	$>$			381.01	$\approx$	377.96
Kagel and Schley (2013)	2709.95	$\ll$	2993.4	$\gg$	2539.99		2539.99	$\approx$			2561.76	$\gg$	2450.24
Sherstyuk et al. (2013)	1322.6	$\ll$	1450	$\gg$	1298.37		1298.37	$\approx$			1303.8	$\approx$	1234.52
Pooled	**25933.42**	$\ll$	**27915.32**	$\gg$	**25504.76**		**25504.76**	$\gg$			**24779.85**	$\gg$	**24202.51**
Second halves per session													
Aoyagi and Frechette (2009)	416.51	$\approx$	437.8	$\approx$	423.05		416.51	$\approx$			423.68	$\approx$	421.21
Blonski et al. (2011)	1414.39	$\ll$	1553.12	$\approx$	1609.79		1414.39	$>$			1346.79	$\approx$	1370.16
Bruttel and Kamecke (2012)	538.17	$<$	611.91	$\gg$	525.5		538.17	$\approx$			536.77	$\gg$	480.47
Dal Bó (2005)	737.05	$<$	786.21	$>$	741.54		737.05	$>$			699.05	$\approx$	677.24
Dal Bó and Fréchette (2011)	5220.17	$\ll$	6378.16	$\gg$	5069.04		5220.17	$\approx$			5128.69	$\gg$	4565.93
Dal Bó and Fréchette (2015)	8205.77	$\ll$	9401.19	$\gg$	7947.33		8205.77	$\gg$			7825.98	$\gg$	7306.25
Dreber et al. (2008)	619.9	$\approx$	662.24	$>$	596.78		619.9	$\approx$			589.84	$>$	544.66
Duffy and Ochs (2009)	1883.52	$\approx$	1914.83	$>$	1850.35		1883.52	$>$			1761.6	$\gg$	1656.55
Fréchette and Yuksel (2017)	438.55	$<$	478.2	$\approx$	434.61		438.55	$\approx$			423.34	$\approx$	422.61
Fudenberg et al. (2012)	514.87	$\approx$	516.12	$\approx$	515.97		514.87	$\approx$			452.6	$\approx$	433.74
Kagel and Schley (2013)	1808.21	$\ll$	2336.29	$\gg$	1718.07		1808.21	$\approx$			1775.62	$\gg$	1572.95
Sherstyuk et al. (2013)	955.73	$\ll$	1137.49	$\gg$	958.99		955.73	$\approx$			951.34	$\gg$	834.74
Pooled	**23009.84**	$\ll$	**26479.73**	$\gg$	**23143.38**		**23009.84**	$\gg$			**22097.67**	$\gg$	**20541.83**

Note: This table extends Table 21 by picking the best switching model per half-session, after picking the best-fitting mixture involving the generalized forms of AD, Grim, TFT, AC and WSLS (as above) for each treatment independently, and examining its goodness-of-fit in relation to Semi-Grim and mixtures involving Semi-Grim.

Table 26: Table 25 by treatments – Best mixtures of generalized strategies in relation to a Semi-Grim strategy

(a) First halves per session

	Best mixture of generalized strategies											
	No Switching		Random Switching		Markov Switching		Best Switching		Semi-Grim		AD + SG	
Specification												
# Models evaluated	8^{32}		8^{32}		8^{32}				1		1	
# Pars estimated (by treatment)	64		64		196				3		5	
# Parameters accounted for	6–10		6–10		15-35				3		5	
AF09–34	756.95	≈	763.11	≈	755.97		755.97	≈	781.86	≈	792.51	
BOS11–9	89.7	≈	87.81	≈	91.36		91.36	≈	86.56	≈	88.35	
BOS11–14	102.27	≈	94.44	≈	96.5		96.5	≈	93.88	≈	98.01	
BOS11–15	38.79	≈	37.18	≈	44.74		44.74	≈	37.73	<	43.07	
BOS11–16	168.92	≈	176.43	≈	174.73		174.73	≈	167.42	≈	157.13	
BOS11–17	115.07	≈	123.19	≈	123.74		123.74	>	115.02	≈	119.79	
BOS11–26	257.82	≈	269.17	≈	256.37		256.37	≈	244.5	≈	246.46	
BOS11–27	103.44	≈	114.9	≈	110.81		110.81	>	92.83	≈	92.07	
BOS11–30	60.42	≈	68.55	≈	68.16		68.16	≫	55.74	<	61.12	
BOS11–31	129.65	≈	137.48	≈	145.13		145.13	≈	125.52	≈	128.49	
BK12–28	817.89	≈	835.6	>	787.63		787.63	≈	800.12	≈	771.14	
D05–18	241.44	≈	230.66	≈	238.66		238.66	≈	230.57	≈	238.35	
D05–19	396.28	≪	439.65	>	403.81		403.81	≈	395.06	≈	375.07	
DF11–6	823.69	≈	909.31	>	772.55		772.55	≈	794.44	≈	748.9	
DF11–7	1297.64	<	1370.65	≫	1181.36		1181.36	<	1256.83	≈	1229.61	
DF11–8	1422.73	≪	1668.83	≫	1284.25		1284.25	<	1389.06	≈	1286.68	
DF11–22	1080.23	≈	1110.68	≈	1056.77		1056.77	>	965.96	≈	961.58	
DF11–23	1082.68	<	1185.69	≫	1027.3		1027.3	≈	1019.57	≈	972.17	
DF11–24	1171.57	≈	1179.6	≫	1022.62		1022.62	≪	1145.15	≈	1115.95	
DF15–4	439.54	≈	474.37	≈	412.27		412.27	≈	436.57	≈	422.43	
DF15–5	1762.23	≪	2211.09	≫	1638.92		1638.92	≪	1738.77	≫	1649.1	
DF15–20	1463.03	<	1547.14	≫	1433.87		1433.87	≈	1405.3	>	1366.09	
DF15–21	1974.94	≪	2184.97	≫	1902.95		1902.95	≈	1872.47	>	1827.37	
DF15–33	2379.17	≈	2350.87	≈	2296.41		2296.41	>	2186.26	≈	2178.12	
DF15–35	384.6	≈	372.62	≈	382.07		382.07	>	349.06	≈	346.18	
DRFN08–10	374.3	≈	391.56	>	339.73		339.73	≈	367.86	≈	359.04	
DRFN08–11	408.63	<	468.48	≫	413.19		413.19	≈	411.01	≈	400.49	
DO09–32	1395.4	<	1461.01	>	1394.31		1394.31	≈	1372.97	≈	1361.15	
FY17–25	300.87	<	345.74	>	298.53		298.53	≈	299.62	≈	289.54	
FRD12–29	432.32	≈	432.38	≈	425.54		425.54	>	381.01	≈	377.96	
KS13–12	2709.95	≪	2993.4	≫	2539.99		2539.99	≈	2561.76	≫	2450.24	
STS13–13	1322.6	≪	1450	≫	1298.37		1298.37	≈	1303.8	≈	1234.52	
Aoyagi and Frechette (2009)	756.95	≈	763.11	≈	755.97		755.97	≈	781.86	≈	792.51	
Blonski et al. (2011)	1134.67	≈	1173.15	≪	1272.13		1272.13	≫	1069.28	≈	1104.6	
Bruttel and Kamecke (2012)	817.89	≈	835.6	>	787.63		787.63	≈	800.12	≈	771.14	
Dal Bó (2005)	641.98	<	674.57	≈	653.11		653.11	≈	629.17	≈	618.39	
Dal Bó and Fréchette (2011)	6921.58	≪	7467.72	≫	6465.99		6465.99	≈	6597.93	>	6352.59	
Dal Bó and Fréchette (2015)	8446	≪	9183.55	≫	8168.2		8168.2	>	8017.59	≫	7830.12	
Dreber et al. (2008)	787.71	<	865.64	≫	763.43		763.43	≈	782.37	≈	764.44	
Duffy and Ochs (2009)	1395.4	<	1461.01	>	1394.31		1394.31	≈	1372.97	≈	1361.15	
Fréchette and Yuksel (2017)	300.87	<	345.74	>	298.53		298.53	≈	299.62	≈	289.54	
Fudenberg et al. (2012)	432.32	≈	432.38	≈	425.54		425.54	>	381.01	≈	377.96	
Kagel and Schley (2013)	2709.95	≪	2993.4	≫	2539.99		2539.99	≈	2561.76	≫	2450.24	
Sherstyuk et al. (2013)	1322.6	≪	1450	≫	1298.37		1298.37	≈	1303.8	≈	1234.52	
Pooled	**25933.42**	≪	**27915.32**	≫	**25504.76**		**25504.76**	≫	**24779.85**	≫	**24202.51**	

(b) Second halves per session

	Best mixture of generalized strategies											
	No Switching		Random Switching		Markov Switching		Best Switching		Semi-Grim		AD + SG	
Specification												
# Models evaluated	8^{32}		8^{32}		8^{32}				1		1	
# Pars estimated (by treatment)	64		64		196				3		5	
# Parameters accounted for	6–10		6–10		15-35				3		5	
AF09–34	416.51	≈	437.8	≈	423.05		416.51	≈	423.68	≈	421.21	
BOS11–9	78.84	<	103.47	≈	83.98		78.84	≈	75.1	≈	80.12	
BOS11–14	45.31	≈	42.43	≈	48.97		45.31	≈	35.58	≈	35.86	
BOS11–15	20.01	≈	20.01	≪	33.5		20.01	≈	19.23	<	24.71	
BOS11–16	148.98	≈	168.12	≈	158.84		148.98	≈	150.95	≈	138.89	
BOS11–17	211.59	≈	225.1	≈	216.6		211.59	≈	196.25	≈	201.03	
BOS11–26	327.16	≈	352.05	≈	338.09		327.16	≈	299.85	≈	309.63	
BOS11–27	224.85	≈	254.56	≈	233.57		224.85	≈	235.88	≈	223.91	
BOS11–30	139.46	≈	139.47	≈	146.9		139.46	≈	129.86	≈	132.45	
BOS11–31	151.87	<	179.31	≈	171.15		151.87	≈	154.02	≈	153.45	
BK12–28	538.17	<	611.91	≫	525.5		538.17	≈	536.77	≫	480.47	
D05–18	340.33	≈	355.81	≈	346.45		340.33	≈	334.18	>	312.74	
D05–19	392.47	<	426.15	>	384.46		392.47	>	361.33	≈	359.54	
DF11–6	579.84	≈	628.84	≈	565.43		579.84	≈	526.15	≈	489.74	
DF11–7	1359.89	≪	1582.11	≫	1299.86		1359.89	≈	1316.79	≈	1250.02	
DF11–8	1028.93	≪	1600.86	≫	904.89		1028.93	≈	1078.24	>	871.84	
DF11–22	1012.26	<	1102.07	≫	973.62		1012.26	≈	930.3	>	858.03	
DF11–23	743.89	<	943.35	≫	739.29		743.89	≈	767.04	>	608.87	
DF11–24	450.61	≈	477.85	≈	455.62		450.61	≈	483.25	≈	449.72	
DF15–4	301.69	<	385.5	≈	307.03		301.69	≈	320.02	≈	299.49	
DF15–5	1581.28	≪	2217.39	≫	1435.63		1581.28	≈	1606.96	≫	1407.26	
DF15–20	1273.14	≪	1441.16	≫	1270.1		1273.14	≈	1232.33	≫	1145.96	
DF15–21	1688.09	<	1878.92	≫	1544.66		1688.09	≈	1591.27	≫	1453.74	
DF15–33	2582.61	<	2690.87	≫	2541.5		2582.61	>	2405.23	≈	2331.38	
DF15–35	733.28	≈	742.93	≈	723.78		733.28	≫	641.02	≈	627.6	
DRFN08–10	276.61	≈	285.26	≈	243.71		276.61	>	244.91	≈	234.76	
DRFN08–11	339.09	≈	371.38	≈	339.95		339.09	≈	341.42	>	305	
DO09–32	1883.52	≈	1914.83	>	1850.35		1883.52	>	1761.6	≫	1656.55	
FY17–25	438.55	<	478.2	≈	434.61		438.55	≈	423.34	≈	422.61	
FRD12–29	514.87	≈	516.12	≈	515.97		514.87	≈	452.6	≈	433.74	
KS13–12	1808.21	≪	2336.29	≫	1718.07		1808.21	≈	1775.62	≫	1572.95	
STS13–13	955.73	≪	1137.49	≫	958.99		955.73	≈	951.34	≫	834.74	
Aoyagi and Frechette (2009)	416.51	≈	437.8	≈	423.05		416.51	≈	423.68	≈	421.21	
Blonski et al. (2011)	1414.39	≪	1553.12	≈	1609.79		1414.39	>	1346.79	≈	1370.16	
Bruttel and Kamecke (2012)	538.17	<	611.91	≫	525.5		538.17	≈	536.77	≫	480.47	
Dal Bó (2005)	737.05	<	786.21	>	741.54		737.05	>	699.05	≈	677.24	
Dal Bó and Fréchette (2011)	5220.17	≪	6378.16	≫	5069.04		5220.17	≈	5128.69	≫	4565.93	
Dal Bó and Fréchette (2015)	8205.77	≪	9401.19	≫	7947.33		8205.77	≫	7825.98	≫	7306.25	
Dreber et al. (2008)	619.9	≈	662.24	>	596.78		619.9	≈	589.84	>	544.66	
Duffy and Ochs (2009)	1883.52	≈	1914.83	>	1850.35		1883.52	>	1761.6	≫	1656.55	
Fréchette and Yuksel (2017)	438.55	<	478.2	≈	434.61		438.55	≈	423.34	≈	422.61	
Fudenberg et al. (2012)	514.87	≈	516.12	≈	515.97		514.87	≈	452.6	≈	433.74	
Kagel and Schley (2013)	1808.21	≪	2336.29	≫	1718.07		1808.21	≈	1775.62	≫	1572.95	
Sherstyuk et al. (2013)	955.73	≪	1137.49	≫	958.99		955.73	≈	951.34	≫	834.74	
Pooled	**23009.84**	≪	**26479.73**	≫	**23143.38**		**23009.84**	≫	**22097.67**	≫	**20541.83**	

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 14.

Table 27: Best mixtures of pure or generalized strategies in relation to Semi-Grim
(ICL-BIC of the models, less is better and relation signs point toward better models)

			Best mixture of pure or generalized strategies										Best Mixture	
	Baseline Model		No Switching		Random Switching		Markov Switching	Best Switching		AD + SG		AD + 2SG		Best Switching By Treatment
Specification														
# Models evaluated	1		13^{32}		13^{32}		13^{32}	3×13^{32}		1		1		$39^{32} \approx 10^{51}$
# Pars estimated (by treatment)	5		80		80		278	438		5		9		438
# Parameters accounted for	5		3–10		3–10		12-35	3–30		5		9		3–30
First halves per session														
Aoyagi and Frechette (2009)	886.44	$\gg$	756.95	$\approx$	763.11	$\approx$	755.97	755.97	$\approx$	781.86	$\approx$	777.3	$\approx$	755.97
Blonski et al. (2011)	1114.69	$\gg$	1069.58	$\approx$	1104.85	$\ll$	1225.35	1225.35	$\gg$	1069.28	$<$	1134.96	$>$	1069.39
Bruttel and Kamecke (2012)	845.41	$\approx$	817.89	$\approx$	835.6	$>$	785.49	785.49	$\approx$	800.12	$>$	762.83	$\approx$	785.49
Dal Bó (2005)	666.1	$>$	635.04	$<$	674.57	$\approx$	648.75	648.75	$\approx$	629.17	$\approx$	600.26	$<$	631.2
Dal Bó and Fréchette (2011)	7423.23	$\gg$	6904.79	$\ll$	7456.12	$\gg$	6388.49	6388.49	$<$	6597.93	$\gg$	6304.97	$\approx$	6388.49
Dal Bó and Fréchette (2015)	8880.62	$\gg$	8434.93	$\ll$	9166.72	$\gg$	8158.31	8158.31	$>$	8017.59	$\gg$	7810.7	$\ll$	8138.61
Dreber et al. (2008)	871.32	$\gg$	787.71	$<$	863.7	$\gg$	752.16	752.16	$\approx$	782.37	$\approx$	763.52	$\approx$	752.16
Duffy and Ochs (2009)	1448.71	$\approx$	1395.4	$<$	1461.01	$>$	1372.99	1372.99	$\approx$	1372.97	$\approx$	1320.71	$\approx$	1372.99
Fréchette and Yuksel (2017)	321.32	$\approx$	300.87	$<$	337.5	$>$	298.53	298.53	$\approx$	299.62	$\approx$	284.11	$\approx$	298.53
Fudenberg et al. (2012)	454.09	$\approx$	432.32	$\approx$	432.38	$\approx$	425.54	425.54	$>$	381.01	$\approx$	370.01	$<$	425.54
Kagel and Schley (2013)	2735.02	$\approx$	2685.4	$\ll$	2993.4	$\gg$	2439.06	2439.06	$\approx$	2561.76	$\gg$	2421.27	$\approx$	2439.06
Sherstyuk et al. (2013)	1389.33	$\approx$	1322.6	$\ll$	1450	$\gg$	1296.85	1296.85	$\approx$	1303.8	$\gg$	1200.28	$<$	1296.85
Pooled	**27218.66**	$\gg$	**25758.38**	$\ll$	**27754.81**	$\gg$	**25166.24**	**25166.24**	$>$	**24779.85**	$\gg$	**24079.18**	$\ll$	**24863.15**
Second halves per session														
Aoyagi and Frechette (2009)	534.29	$\gg$	416.51	$\approx$	437.8	$\approx$	423.05	416.51	$\approx$	423.68	$\approx$	422.24	$\approx$	416.51
Blonski et al. (2011)	1503.41	$\gg$	1398.5	$\ll$	1509.09	$<$	1593.01	1398.5	$>$	1346.79	$\approx$	1385.91	$\approx$	1394.16
Bruttel and Kamecke (2012)	588.33	$>$	538.17	$<$	611.91	$\gg$	516.71	538.17	$\approx$	536.77	$>$	478.23	$\approx$	516.71
Dal Bó (2005)	751.82	$\approx$	732.27	$<$	786.21	$>$	739.59	732.27	$>$	699.05	$\approx$	679.04	$<$	729.48
Dal Bó and Fréchette (2011)	6065.93	$\gg$	5195.88	$\ll$	6378.16	$\gg$	5007.24	5195.88	$\approx$	5128.69	$\gg$	4545.08	$\ll$	4964.77
Dal Bó and Fréchette (2015)	9085.4	$\gg$	8177.46	$\ll$	9401.19	$\gg$	7910.83	8177.46	$\gg$	7825.98	$\gg$	7310.27	$\ll$	7893.79
Dreber et al. (2008)	656.38	$\approx$	619.9	$\approx$	662.24	$>$	581.94	619.9	$\approx$	589.84	$>$	541.83	$\approx$	581.94
Duffy and Ochs (2009)	2010.01	$>$	1883.52	$\approx$	1914.83	$>$	1850.35	1883.52	$>$	1761.6	$>$	1602.93	$\ll$	1850.35
Fréchette and Yuksel (2017)	469.85	$\approx$	433.18	$<$	474.93	$>$	427.79	433.18	$\approx$	423.34	$\approx$	381.63	$\ll$	427.79
Fudenberg et al. (2012)	530.3	$\approx$	514.87	$\approx$	516.12	$\approx$	515.97	514.87	$\approx$	452.6	$>$	414.24	$<$	514.87
Kagel and Schley (2013)	1866.19	$\approx$	1751.81	$\ll$	2336.29	$\gg$	1678.7	1751.81	$\approx$	1775.62	$\gg$	1541.38	$<$	1678.7
Sherstyuk et al. (2013)	1027.43	$>$	955.73	$\ll$	1137.49	$\gg$	958.99	955.73	$\approx$	951.34	$\gg$	823.06	$\ll$	955.73
Pooled	**25271.72**	$\gg$	**22848.49**	$\ll$	**26409.44**	$\gg$	**22927.9**	**22848.49**	$\gg$	**22097.67**	$\gg$	**20454.13**	$\ll$	**22422.07**

Note: This table extends Table 21 by picking the best switching model per half-session, after picking the best-fitting mixture involving either the pure or generalized forms of AD, Grim, TFT, AC and WSLS (as above) for each treatment independently, and examining its goodness-of-fit in relation to Semi-Grim and mixtures involving Semi-Grim. The model "AD+SG2" has the same number of degrees of freedom as the Semi-Grim model.

Table 28: Table 27 by treatments – Best mixtures of pure or generalized strategies in relation to Semi-Grim

(a) First halves per session

	Baseline Model		Best mixture of pure or generalized strategies: No Switching		Random Switching		Markov Switching	Best Switching		AD + SG		AD + 2SG	Best Mixture: Best Switching By Treatment
Specification													
# Models evaluated	1		13^{32}		13^{32}		13^{32}	3×13^{32}		1		1	$39^{32} \approx 10^{51}$
# Pars estimated (by treatment)	5		80		80		278	438		5		9	438
# Parameters accounted for	5		3–10		3–10		12-35	3–30		5		9	3–30
AF09–34	886.44	$\gg$	756.95	$\approx$	763.11	$\approx$	755.97	755.97	$\approx$	781.86	$\approx$	777.3	$\approx$	755.97
BOS11–9	85.17	$\approx$	83.42	$\approx$	83.96	$\approx$	88.41	88.41	$\approx$	86.56	$\approx$	85.71	$\approx$	83.42
BOS11–14	100.72	$\approx$	97.73	$\approx$	90	$\approx$	92.94	92.94	$\approx$	93.88	$\approx$	95.87	$\approx$	90
BOS11–15	37.29	$\approx$	34.3	$\approx$	32.69	$\ll$	43.18	43.18	$>$	37.73	$\ll$	53.61	$\gg$	32.69
BOS11–16	176.55	$\approx$	167.3	$\approx$	169.38	$\approx$	170.57	170.57	$\approx$	167.42	$\approx$	157.72	$\approx$	167.3
BOS11–17	113.57	$\approx$	110.57	$\approx$	118.71	$\approx$	121.05	121.05	$>$	115.02	$<$	122.41	$>$	110.57
BOS11–26	260.57	$\approx$	256.88	$\approx$	262.33	$\approx$	256.37	256.37	$\approx$	244.5	$\approx$	249.78	$\approx$	256.37
BOS11–27	103.61	$\approx$	102.11	$\approx$	112.76	$\approx$	110.81	110.81	$>$	92.83	$\approx$	93.42	$\approx$	102.11
BOS11–30	59.81	$>$	56.81	$\approx$	65.61	$\approx$	64.33	64.33	$>$	55.74	$\ll$	64.33	$>$	56.81
BOS11–31	127.32	$\approx$	125.82	$\approx$	135.1	$\approx$	142.43	142.43	$\approx$	125.52	$\approx$	121.98	$\approx$	125.82
BK12–28	845.41	$\approx$	817.89	$\approx$	835.6	$>$	785.49	785.49	$\approx$	800.12	$>$	762.83	$\approx$	785.49
D05–18	241.39	$\approx$	235.84	$\approx$	230.66	$\approx$	235.63	235.63	$\approx$	230.57	$\approx$	229.01	$\approx$	230.66
D05–19	421.17	$>$	396.28	$\ll$	439.65	$>$	403.81	403.81	$\approx$	395.06	$>$	364.87	$<$	396.28
DF11–6	880.04	$\approx$	810.5	$\approx$	909.31	$>$	770.36	770.36	$\approx$	794.44	$\approx$	752.56	$\approx$	770.36
DF11–7	1423.93	$>$	1297.64	$\approx$	1364.07	$\gg$	1132.04	1132.04	$<$	1256.83	$\approx$	1238	$>$	1132.04
DF11–8	1515.51	$>$	1422.73	$\ll$	1668.83	$\gg$	1279.8	1279.8	$<$	1389.06	$\approx$	1289.21	$\approx$	1279.8
DF11–22	1192.92	$>$	1080.23	$\approx$	1110.68	$\approx$	1056.77	1056.77	$>$	965.96	$\approx$	944.2	$\ll$	1056.77
DF11–23	1144.78	$\approx$	1082.68	$<$	1185.69	$\gg$	1020.09	1020.09	$\approx$	1019.57	$>$	941.73	$\approx$	1020.09
DF11–24	1239.14	$>$	1171.57	$\approx$	1179.6	$\gg$	1022.62	1022.62	$\ll$	1145.15	$\approx$	1090.82	$\approx$	1022.62
DF15–4	460.23	$>$	431.07	$\approx$	467.36	$>$	395.89	395.89	$\approx$	436.57	$\approx$	425	$>$	395.89
DF15–5	1808.3	$>$	1762.23	$\ll$	2211.09	$\gg$	1638.92	1638.92	$\ll$	1738.77	$\gg$	1639.52	$\approx$	1638.92
DF15–20	1588.62	$\gg$	1463.03	$<$	1543.46	$\gg$	1433.87	1433.87	$\approx$	1405.3	$>$	1364.37	$\approx$	1433.87
DF15–21	2015.1	$>$	1974.94	$\ll$	2184.97	$\gg$	1902.95	1902.95	$\approx$	1872.47	$>$	1811.11	$<$	1902.95
DF15–33	2573.89	$\gg$	2379.17	$\approx$	2350.87	$\approx$	2296.41	2296.41	$>$	2186.26	$\approx$	2174.56	$\ll$	2296.41
DF15–35	405.32	$\approx$	384.6	$\approx$	372.62	$\approx$	382.07	382.07	$>$	349.06	$\approx$	343.65	$<$	372.62
DRFN08–10	413.58	$>$	374.3	$\approx$	390.77	$>$	334.73	334.73	$<$	367.86	$\approx$	358.63	$\approx$	334.73
DRFN08–11	454.24	$>$	408.63	$<$	468.48	$\gg$	405.73	405.73	$\approx$	411.01	$\approx$	398.59	$\approx$	405.73
DO09–32	1448.71	$\approx$	1395.4	$<$	1461.01	$>$	1372.99	1372.99	$\approx$	1372.97	$\approx$	1320.71	$\approx$	1372.99
FY17–25	321.32	$\approx$	300.87	$<$	337.5	$>$	298.53	298.53	$\approx$	299.62	$\approx$	284.11	$\approx$	298.53
FRD12–29	454.09	$\approx$	432.32	$\approx$	432.38	$\approx$	425.54	425.54	$>$	381.01	$\approx$	370.01	$<$	425.54
KS13–12	2735.02	$\approx$	2685.4	$\ll$	2993.4	$\gg$	2439.06	2439.06	$\approx$	2561.76	$\gg$	2421.27	$\approx$	2439.06
STS13–13	1389.33	$\approx$	1322.6	$\ll$	1450	$\gg$	1296.85	1296.85	$\approx$	1303.8	$\gg$	1200.28	$<$	1296.85
Aoyagi and Frechette (2009)	886.44	$\gg$	756.95	$\approx$	763.11	$\approx$	755.97	755.97	$\approx$	781.86	$\approx$	777.3	$\approx$	755.97
Blonski et al. (2011)	1114.69	$\gg$	1069.58	$\approx$	1104.85	$\ll$	1225.35	1225.35	$\gg$	1069.28	$<$	1134.96	$>$	1069.39
Bruttel and Kamecke (2012)	845.41	$\approx$	817.89	$\approx$	835.6	$>$	785.49	785.49	$\approx$	800.12	$>$	762.83	$\approx$	785.49
Dal Bó (2005)	666.1	$>$	635.04	$<$	674.57	$\approx$	648.75	648.75	$\approx$	629.17	$\approx$	600.26	$<$	631.2
Dal Bó and Fréchette (2011)	7423.23	$\gg$	6904.79	$\ll$	7456.12	$\gg$	6388.49	6388.49	$<$	6597.93	$\gg$	6304.97	$\approx$	6388.49
Dal Bó and Fréchette (2015)	8880.62	$\gg$	8434.93	$\ll$	9166.72	$\gg$	8158.31	8158.31	$>$	8017.59	$\gg$	7810.7	$\ll$	8138.61
Dreber et al. (2008)	871.32	$\gg$	787.71	$<$	863.7	$\gg$	752.16	752.16	$\approx$	782.37	$\approx$	763.52	$\approx$	752.16
Duffy and Ochs (2009)	1448.71	$\approx$	1395.4	$<$	1461.01	$>$	1372.99	1372.99	$\approx$	1372.97	$\approx$	1320.71	$\approx$	1372.99
Fréchette and Yuksel (2017)	321.32	$\approx$	300.87	$<$	337.5	$>$	298.53	298.53	$\approx$	299.62	$\approx$	284.11	$\approx$	298.53
Fudenberg et al. (2012)	454.09	$\approx$	432.32	$\approx$	432.38	$\approx$	425.54	425.54	$>$	381.01	$\approx$	370.01	$<$	425.54
Kagel and Schley (2013)	2735.02	$\approx$	2685.4	$\ll$	2993.4	$\gg$	2439.06	2439.06	$\approx$	2561.76	$\gg$	2421.27	$\approx$	2439.06
Sherstyuk et al. (2013)	1389.33	$\approx$	1322.6	$\ll$	1450	$\gg$	1296.85	1296.85	$\approx$	1303.8	$\gg$	1200.28	$<$	1296.85
Pooled	**27218.66**	$\gg$	**25758.38**	$\ll$	**27754.81**	$\gg$	**25166.24**	**25166.24**	$>$	**24779.85**	$\gg$	**24079.18**	$\ll$	**24863.15**

(b) Second halves per session

	Baseline Model		Best mixture of pure or generalized strategies: No Switching		Random Switching		Markov Switching	Best Switching		AD + SG		AD + 2SG	Best Mixture: Best Switching By Treatment
Specification													
# Models evaluated	1		13^{32}		13^{32}		13^{32}	3×13^{32}		1		1	$39^{32} \approx 10^{51}$
# Pars estimated (by treatment)	5		80		80		278	438		5		9	438
# Parameters accounted for	5		3–10		3–10		12-35	3–30		5		9	3–30
AF09–34	534.29	$\gg$	416.51	$\approx$	437.8	$\approx$	423.05	416.51	$\approx$	423.68	$\approx$	422.24	$\approx$	416.51
BOS11–9	87.22	$\approx$	78.84	$\approx$	96.42	$\approx$	83.98	78.84	$\approx$	75.1	$\approx$	76.06	$\approx$	78.84
BOS11–14	43.82	$\approx$	40.82	$\approx$	40.83	$\approx$	48.97	40.82	$\approx$	35.58	$\approx$	38.21	$\approx$	40.82
BOS11–15	18.52	$\approx$	15.52	$\approx$	15.52	$\ll$	29.01	15.52	$\approx$	19.23	$\ll$	30.15	$\gg$	15.52
BOS11–16	160.48	$\approx$	148.98	$\approx$	165.09	$\approx$	157.84	148.98	$\approx$	150.95	$\approx$	149.71	$\approx$	148.98
BOS11–17	232.75	$>$	211.59	$\approx$	225.1	$\approx$	216.6	211.59	$\approx$	196.25	$\approx$	198.59	$\approx$	211.59
BOS11–26	369.98	$\gg$	327.16	$\approx$	352.05	$\approx$	338.09	327.16	$\approx$	299.85	$\approx$	295.25	$<$	327.16
BOS11–27	228.41	$\approx$	224.85	$\approx$	254.56	$\approx$	233.57	224.85	$\approx$	235.88	$\approx$	224.21	$\approx$	224.85
BOS11–30	149.49	$\approx$	139.46	$\approx$	137.43	$\approx$	145.96	139.46	$\approx$	129.86	$\approx$	133.81	$\approx$	137.43
BOS11–31	162.67	$\approx$	151.87	$\approx$	174.2	$\approx$	171.15	151.87	$\approx$	154.02	$\approx$	149.77	$\approx$	151.87
BK12–28	588.33	$>$	538.17	$<$	611.91	$\gg$	516.71	538.17	$\approx$	536.77	$>$	478.23	$\approx$	516.71
D05–18	355.62	$\approx$	340.33	$\approx$	355.81	$\approx$	346.45	340.33	$\approx$	334.18	$\approx$	315.38	$<$	340.33
D05–19	392.65	$\approx$	388.49	$<$	426.15	$>$	383.3	388.49	$>$	361.33	$\approx$	357.27	$<$	383.3
DF11–6	751.56	$>$	579.84	$\approx$	628.84	$\approx$	557.16	579.84	$\approx$	526.15	$\approx$	509.84	$\approx$	557.16
DF11–7	1571.76	$\gg$	1359.89	$\ll$	1582.11	$\gg$	1268.34	1359.89	$\approx$	1316.79	$>$	1246.93	$\approx$	1268.34
DF11–8	1142.1	$>$	1028.93	$\ll$	1600.86	$\gg$	904.89	1028.93	$\approx$	1078.24	$>$	872.45	$\approx$	904.89
DF11–22	1198.53	$>$	1012.26	$<$	1102.07	$\gg$	973.62	1012.26	$\approx$	930.3	$>$	860.39	$<$	973.62
DF11–23	842.37	$\approx$	723.5	$<$	943.35	$>$	737.29	723.5	$\approx$	767.04	$>$	582.24	$<$	723.5
DF11–24	532.68	$\gg$	450.61	$\approx$	477.85	$\approx$	455.62	450.61	$\approx$	483.25	$>$	424.76	$\approx$	450.61
DF15–4	345.97	$>$	301.69	$<$	385.5	$\approx$	307.03	301.69	$\approx$	320.02	$\approx$	306.89	$\approx$	301.69
DF15–5	1686.18	$>$	1581.28	$\ll$	2217.39	$\gg$	1435.63	1581.28	$\approx$	1606.96	$\gg$	1409.52	$\approx$	1435.63
DF15–20	1572.51	$\gg$	1273.14	$\ll$	1441.16	$\gg$	1270.1	1273.14	$\approx$	1232.33	$\gg$	1141.29	$<$	1270.1
DF15–21	1754.13	$\approx$	1664.01	$\ll$	1878.92	$\gg$	1504.63	1664.01	$\approx$	1591.27	$\gg$	1422.87	$\approx$	1504.63
DF15–33	2915.83	$\gg$	2582.61	$<$	2690.87	$\gg$	2541.5	2582.61	$>$	2405.23	$\approx$	2349.48	$\ll$	2541.5
DF15–35	781.64	$>$	733.28	$\approx$	742.93	$\approx$	723.78	733.28	$\gg$	641.02	$\approx$	627.73	$<$	723.78
DRFN08–10	304.41	$\approx$	276.61	$\approx$	285.26	$\approx$	243.71	276.61	$>$	244.91	$\approx$	236.44	$\approx$	243.71
DRFN08–11	348.47	$\approx$	339.09	$\approx$	371.38	$\approx$	323.06	339.09	$\approx$	341.42	$\gg$	299.09	$\approx$	323.06
DO09–32	2010.01	$>$	1883.52	$\approx$	1914.83	$>$	1850.35	1883.52	$>$	1761.6	$>$	1602.93	$\ll$	1850.35
FY17–25	469.85	$\approx$	433.18	$<$	474.93	$>$	427.79	433.18	$\approx$	423.34	$\approx$	381.63	$\ll$	427.79
FRD12–29	530.3	$\approx$	514.87	$\approx$	516.12	$\approx$	515.97	514.87	$\approx$	452.6	$>$	414.24	$<$	514.87
KS13–12	1866.19	$\approx$	1751.81	$\ll$	2336.29	$\gg$	1678.7	1751.81	$\approx$	1775.62	$\gg$	1541.38	$<$	1678.7
STS13–13	1027.43	$>$	955.73	$\ll$	1137.49	$\gg$	958.99	955.73	$\approx$	951.34	$\gg$	823.06	$\ll$	955.73
Aoyagi and Frechette (2009)	534.29	$\gg$	416.51	$\approx$	437.8	$\approx$	423.05	416.51	$\approx$	423.68	$\approx$	422.24	$\approx$	416.51
Blonski et al. (2011)	1503.41	$\gg$	1398.5	$\ll$	1509.09	$<$	1593.01	1398.5	$>$	1346.79	$\approx$	1385.91	$\approx$	1394.16
Bruttel and Kamecke (2012)	588.33	$>$	538.17	$<$	611.91	$\gg$	516.71	538.17	$\approx$	536.77	$>$	478.23	$\approx$	516.71
Dal Bó (2005)	751.82	$\approx$	732.27	$<$	786.21	$>$	739.59	732.27	$>$	699.05	$\approx$	679.04	$<$	729.48
Dal Bó and Fréchette (2011)	6065.93	$\gg$	5195.88	$\ll$	6378.16	$\gg$	5007.24	5195.88	$\approx$	5128.69	$\gg$	4545.08	$\ll$	4964.77
Dal Bó and Fréchette (2015)	9085.4	$\gg$	8177.46	$\ll$	9401.19	$\gg$	7910.83	8177.46	$\gg$	7825.98	$\gg$	7310.27	$\ll$	7893.79
Dreber et al. (2008)	656.38	$\approx$	619.9	$\approx$	662.24	$>$	581.94	619.9	$\approx$	589.84	$>$	541.83	$\approx$	581.94
Duffy and Ochs (2009)	2010.01	$>$	1883.52	$\approx$	1914.83	$>$	1850.35	1883.52	$>$	1761.6	$>$	1602.93	$\ll$	1850.35
Fréchette and Yuksel (2017)	469.85	$\approx$	433.18	$<$	474.93	$>$	427.79	433.18	$\approx$	423.34	$\approx$	381.63	$\ll$	427.79
Fudenberg et al. (2012)	530.3	$\approx$	514.87	$\approx$	516.12	$\approx$	515.97	514.87	$\approx$	452.6	$>$	414.24	$<$	514.87
Kagel and Schley (2013)	1866.19	$\approx$	1751.81	$\ll$	2336.29	$\gg$	1678.7	1751.81	$\approx$	1775.62	$\gg$	1541.38	$<$	1678.7
Sherstyuk et al. (2013)	1027.43	$>$	955.73	$\ll$	1137.49	$\gg$	958.99	955.73	$\approx$	951.34	$\gg$	823.06	$\ll$	955.73
Pooled	**25271.72**	$\gg$	**22848.49**	$\ll$	**26409.44**	$\gg$	**22927.9**	**22848.49**	$\gg$	**22097.67**	$\gg$	**20454.13**	$\ll$	**22422.07**

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 14.

Table 29: **Memory-1 or Memory-2, and semi-grim, pure or generalized pure?** Strategy mixtures are estimated treatment-by-treatment. The resulting ICL-BICs are pooled within experiments and overall (less is better, relation signs point to better models)

	Memory-2 Generalizations of Semi-Grim + AD							Best Mixtures of Generalized Pure Strategies						Best Pure
	M2"General"		M2"TFT"		M2"Grim"		AD+SG	M1+M2"TFT"		M1+M2"Grim"		M1		M1 & M2
Specification														
# Models evaluated	1		1		1		1	22^{32}		22^{32}		13^{32}		5^{32}
# Pars estimated (by treatment)	12		6		6		5	160		160		80		32
# Parameters accounted for	12		6		6		5	6–15		6–15		6–10		3–8
First halves per session														
Aoyagi and Frechette (2009)	756.04	$\approx$	764.13	$\approx$	749.99	$\approx$	781.86	$\approx$	756.95	$\approx$	756.95	$\approx$	756.95	$\ll$ 884.86
Blonski et al. (2011)	1244.76	$\gg$	1121.17	$\approx$	1120.87	$\gg$	1069.28	$\approx$	1069.56	$\approx$	1069.56	$\approx$	1069.58	$\approx$ 1105.98
Bruttel and Kamecke (2012)	807.47	$\approx$	802.89	$\approx$	804.16	$\approx$	800.12	$\approx$	817.89	$\approx$	817.89	$\approx$	817.89	$\approx$ 839.97
Dal Bó (2005)	660.68	$>$	641.34	$\approx$	642.26	$\approx$	629.17	$\approx$	635.04	$\approx$	635.04	$\approx$	635.04	$\approx$ 653.05
Dal Bó and Fréchette (2011)	6671.28	$\approx$	6616.44	$\approx$	6604.7	$\approx$	6597.93	$\ll$	6904.79	$\approx$	6904.79	$\approx$	6904.79	$\ll$ 7391.89
Dal Bó and Fréchette (2015)	8068.37	$\approx$	8028.83	$\approx$	8031.59	$\approx$	8017.59	$\ll$	8423.8	$\approx$	8431.51	$\approx$	8434.93	$\ll$ 8893.78
Dreber et al. (2008)	805.74	$>$	785.48	$\approx$	785.6	$\approx$	782.37	$\approx$	787.71	$\approx$	787.71	$\approx$	787.71	$<$ 863.47
Duffy and Ochs (2009)	1361.84	$\approx$	1377.17	$\approx$	1369.86	$\approx$	1372.97	$\approx$	1395.4	$\approx$	1395.4	$\approx$	1395.4	$\approx$ 1426.34
Fréchette and Yuksel (2017)	305.9	$\approx$	299.72	$\approx$	296.93	$\approx$	299.62	$\approx$	300.87	$\approx$	300.87	$\approx$	300.87	$\approx$ 317.35
Fudenberg et al. (2012)	387.8	$\approx$	379.84	$\approx$	378.07	$\approx$	381.01	$<$	432.32	$\approx$	432.32	$\approx$	432.32	$\approx$ 463.4
Kagel and Schley (2013)	2542.02	$\approx$	2556.45	$\approx$	2552.09	$\approx$	2561.76	$\approx$	2679.23	$\approx$	2685.4	$\approx$	2685.4	$\approx$ 2730.66
Sherstyuk et al. (2013)	1311.64	$\approx$	1307.45	$\approx$	1303.94	$\approx$	1303.8	$\approx$	1322.6	$\approx$	1322.6	$\approx$	1322.6	$<$ 1398.69
Pooled	**25434.21**	$\gg$	**24972.71**	$\approx$	**24931.86**	$\approx$	**24779.85**	$\ll$	**25750.84**	$\approx$	**25757.44**	$\approx$	**25758.38**	$\ll$ **27115.39**
Second halves per session														
Aoyagi and Frechette (2009)	415.47	$\approx$	421.18	$>$	409.19	$\approx$	423.68	$\approx$	416.51	$\approx$	416.51	$\approx$	416.51	$\ll$ 540.47
Blonski et al. (2011)	1518.54	$\gg$	1395.94	$\approx$	1393.41	$\gg$	1346.79	$\approx$	1398.5	$\approx$	1398.5	$\approx$	1398.5	$<$ 1564.48
Bruttel and Kamecke (2012)	536.19	$\approx$	532.08	$\approx$	529.47	$\approx$	536.77	$\approx$	538.17	$\approx$	538.17	$\approx$	538.17	$\approx$ 567.99
Dal Bó (2005)	727.25	$\approx$	710.88	$\approx$	708.32	$\approx$	699.05	$\approx$	726.04	$\approx$	731.81	$\approx$	732.27	$\approx$ 741.2
Dal Bó and Fréchette (2011)	5201.05	$\approx$	5137.82	$\approx$	5132.96	$\approx$	5128.69	$\approx$	5195.88	$\approx$	5195.88	$\approx$	5195.88	$\ll$ 5960.78
Dal Bó and Fréchette (2015)	7840.87	$\approx$	7829.51	$\approx$	7808.63	$\approx$	7825.98	$\ll$	8172.63	$\approx$	8177.46	$\approx$	8177.46	$\ll$ 9143.98
Dreber et al. (2008)	597.17	$\approx$	580.63	$\approx$	570.33	$\approx$	589.84	$\approx$	618.5	$\approx$	618.89	$\approx$	619.9	$\approx$ 648.55
Duffy and Ochs (2009)	1706.1	$\approx$	1753.41	$\approx$	1719.86	$\approx$	1761.6	$\approx$	1857.06	$\approx$	1876.72	$\approx$	1883.52	$\approx$ 2003.41
Fréchette and Yuksel (2017)	422.32	$\approx$	424.41	$\approx$	419.44	$\approx$	423.34	$\approx$	433.18	$\approx$	433.18	$\approx$	433.18	$<$ 464.23
Fudenberg et al. (2012)	452.64	$\approx$	450.08	$\approx$	447.25	$\approx$	452.6	$<$	484.5	$\approx$	477.91	$\approx$	514.87	$\approx$ 534.47
Kagel and Schley (2013)	1782.43	$\approx$	1777.83	$\approx$	1773.55	$\approx$	1775.62	$\approx$	1751.81	$\approx$	1751.81	$\approx$	1751.81	$\approx$ 1830.26
Sherstyuk et al. (2013)	959.21	$\approx$	952.56	$\approx$	949.46	$\approx$	951.34	$\approx$	955.73	$\approx$	955.73	$\approx$	955.73	$\approx$ 1023.43
Pooled	**22669.91**	$\gg$	**22258.14**	$\approx$	**22153.69**	$\approx$	**22097.67**	$\ll$	**22811.34**	$\approx$	**22828.13**	$\approx$	**22848.49**	$\ll$ **25177.57**

Note: Results treatment-by-treatment are in the appendix. The main body contains ICL-BICs aggregated at paper level. Relation signs and p-values are exactly as above, see Table 3. "M2" ("M1") denotes strategies, whose actions may depend on actions in $t-2$ and $t-1$ ($t-1$ only). The supplements "General", "TFT", "Grim" indicate whether parameters of behavior strategies may depend on: all four possible histories in $t-2$ (M2 "General"), whether the opponent cooperated in $t-2$ (M2 "TFT"), or whether there was joint cooperation in $t-2$ (M2 "Grim"). Pure M2 strategies do not have such free parameters. Columns 1-3 contain one memory-2 version of semi-grim each. Column 4 is memory-1 semi-grim. Columns 5-7 are memory-2 and memory-1 versions of generalized prototypical strategies. The last column contains the best fitting combinations of a set of pure memory-1 and memory-2 strategies from the literature (TFT, Grim, AD, Grim2, TF2T, T2, 2TFT, 2PTFT) for definitions see Table 12 in the Online Appendix.

Table 30: Table 8 by treatments – Memory-1 or Memory-2, and semi-grim, pure or generalized pure?

(a) First halves per session

	Memory-2 Generalizations of Semi-Grim + AD						Best Mixtures of Generalized Pure Strategies						Best Pure	
	M2"General"		M2"TFT"		M2"Grim"	AD+SG	M1+M2"TFT"		M1+M2"Grim"		M1		M1 & M2	
Specification														
# Models evaluated	1		1		1	1	22^{32}		22^{32}		13^{32}		5^{32}	
# Pars estimated (by treatment)	12		6		6	5	160		160		80		32	
# Parameters accounted for	12		6		6	5	6–15		6–15		6–10		3–8	
AF09–34	756.04	≈	764.13	≈	749.99	781.86	756.95	≈	756.95	≈	756.95	≪	884.86	
BOS11–9	91.44	≈	87.24	≈	90.13	86.56	83.42	≈	83.42	≈	83.42	≈	85.2	
BOS11–14	103.75	≈	98.34	≈	97.69	93.88	97.73	≈	97.73	≈	97.73	≈	97.73	
BOS11–15	50.59	≫	41.48	≈	41.64	37.73	34.3	≈	34.3	≈	34.3	≈	34.3	
BOS11–16	175.94	>	168.96	≈	168.84	167.42	167.3	≈	167.3	≈	167.3	≈	174.26	
BOS11–17	128.36	≫	118.65	≈	119.18	115.02	110.57	≈	110.57	≈	110.57	≈	110.57	
BOS11–26	253.27	≈	244.2	≈	242.92	244.5	256.88	≈	256.88	≈	256.88	≈	256.88	
BOS11–27	100.21	>	94.88	≈	93.66	92.83	100.97	≈	100.97	≈	102.11	≈	100.97	
BOS11–30	69.22	≫	60.24	≈	60.24	55.74	56.77	≈	56.77	≈	56.81	≈	56.77	
BOS11–31	131.77	≈	127.07	≈	126.46	125.52	125.82	≈	125.82	≈	125.82	≈	156.95	
BK12–28	807.47	≈	802.89	≈	804.16	800.12	817.89	≈	817.89	≈	817.89	≈	839.97	
D05–18	241.79	≈	235.57	≈	236.43	230.57	235.84	≈	235.84	≈	235.84	≈	235.84	
D05–19	408.97	≈	400.1	≈	400.16	395.06	396.28	≈	396.28	≈	396.28	≈	415.08	
DF11–6	806.97	≈	797.22	≈	798.57	794.44	810.5	≈	810.5	≈	810.5	<	877.78	
DF11–7	1252.93	≈	1248.87	≈	1249.66	1256.83	1297.64	≈	1297.64	≈	1297.64	<	1424.78	
DF11–8	1403.92	≈	1393.05	≈	1393.09	1389.06	1422.73	≈	1422.73	≈	1422.73	<	1501.88	
DF11–22	973.99	≈	965.39	≈	969.46	965.96	1080.23	≈	1080.23	≈	1080.23	≪	1188.65	
DF11–23	1026.82	≈	1024.62	≈	1022.57	1019.57	1082.68	≈	1082.68	≈	1082.68	≈	1148.16	
DF11–24	1131.28	≈	1144.21	≈	1128.27	1145.15	1171.57	≈	1171.57	≈	1171.57	<	1224.49	
DF15–4	442.4	≈	439.96	≈	437.41	436.57	431.07	≈	431.07	≈	431.07	≈	456.32	
DF15–5	1752.09	≈	1739.69	≈	1739.11	1738.77	1751.2	≈	1756.46	≈	1762.23	≈	1817.09	
DF15–20	1408.72	≈	1403.6	≈	1408.86	1405.3	1463.03	≈	1463.03	≈	1463.03	<	1585.91	
DF15–21	1871.42	≈	1871.98	≈	1864.48	1872.47	1971.78	≈	1974.94	≈	1974.94	≈	2022.58	
DF15–33	2154.98	≈	2176.37	≈	2184.09	2186.26	2379.17	≈	2379.17	≈	2379.17	≪	2575.64	
DF15–35	357.11	≈	350.57	≈	350.98	349.06	384.6	≈	384.6	≈	384.6	≈	411.07	
DRFN08–10	375.03	≈	366.4	≈	367.62	367.86	374.3	≈	374.3	≈	374.3	≈	410.24	
DRFN08–11	420.9	≈	413.48	≈	412.38	411.01	408.63	≈	408.63	≈	408.63	≈	451.13	
DO09–32	1361.84	≈	1377.17	≈	1369.86	1372.97	1395.4	≈	1395.4	≈	1395.4	≈	1426.34	
FY17–25	305.9	≈	299.72	≈	296.93	299.62	300.87	≈	300.87	≈	300.87	≈	317.35	
FRD12–29	387.8	≈	379.84	≈	378.07	381.01	432.32	≈	432.32	≈	432.32	≈	463.4	
KS13–12	2542.02	≈	2556.45	≈	2552.09	2561.76	2679.23	≈	2685.4	≈	2685.4	≈	2730.66	
STS13–13	1311.64	≈	1307.45	≈	1303.94	1303.8	1322.6	≈	1322.6	≈	1322.6	<	1398.69	
Aoyagi and Frechette (2009)	756.04	≈	764.13	≈	749.99	781.86	756.95	≈	756.95	≈	756.95	≪	884.86	
Blonski et al. (2011)	1244.76	≫	1121.17	≈	1120.87	1069.28	1069.56	≈	1069.56	≈	1069.58	≈	1105.98	
Bruttel and Kamecke (2012)	807.47	≈	802.89	≈	804.16	800.12	817.89	≈	817.89	≈	817.89	≈	839.97	
Dal Bó (2005)	660.68	>	641.34	≈	642.26	629.17	635.04	≈	635.04	≈	635.04	≈	653.05	
Dal Bó and Fréchette (2011)	6671.28	≈	6616.44	≈	6604.7	6597.93	6904.79	≈	6904.79	≈	6904.79	≪	7391.89	
Dal Bó and Fréchette (2015)	8068.37	≈	8028.83	≈	8031.59	8017.59	8423.8	≈	8431.51	≈	8434.93	≪	8893.78	
Dreber et al. (2008)	805.74	>	785.48	≈	785.6	782.37	787.71	≈	787.71	≈	787.71	<	863.47	
Duffy and Ochs (2009)	1361.84	≈	1377.17	≈	1369.86	1372.97	1395.4	≈	1395.4	≈	1395.4	≈	1426.34	
Fréchette and Yuksel (2017)	305.9	≈	299.72	≈	296.93	299.62	300.87	≈	300.87	≈	300.87	≈	317.35	
Fudenberg et al. (2012)	387.8	≈	379.84	≈	378.07	381.01	432.32	≈	432.32	≈	432.32	≈	463.4	
Kagel and Schley (2013)	2542.02	≈	2556.45	≈	2552.09	2561.76	2679.23	≈	2685.4	≈	2685.4	≈	2730.66	
Sherstyuk et al. (2013)	1311.64	≈	1307.45	≈	1303.94	1303.8	1322.6	≈	1322.6	≈	1322.6	<	1398.69	
Pooled	**25434.21**	≫	**24972.71**	≈	**24931.86**	**24779.85**	**25750.84**	≈	**25757.44**	≈	**25758.38**	≪	**27115.39**	

(b) Second halves per session

	Memory-2 Generalizations of Semi-Grim + AD						Best Mixtures of Generalized Pure Strategies						Best Pure	
	M2"General"		M2"TFT"		M2"Grim"	AD+SG	M1+M2"TFT"		M1+M2"Grim"		M1		M1 & M2	
Specification														
# Models evaluated	1		1		1	1	22^{32}		22^{32}		13^{32}		5^{32}	
# Pars estimated (by treatment)	12		6		6	5	160		160		80		32	
# Parameters accounted for	12		6		6	5	6–15		6–15		6–10		3–8	
AF09–34	415.47	≈	421.18	>	409.19	423.68	416.51	≈	416.51	≈	416.51	≪	540.47	
BOS11–9	88.57	>	79.59	≈	79.59	75.1	78.84	≈	78.84	≈	78.84	≈	84.22	
BOS11–14	58.78	≫	40.08	≈	39.03	35.58	40.82	≈	40.82	≈	40.82	≈	40.82	
BOS11–15	33.32	≫	19.62	≈	19.63	19.23	15.52	≈	15.52	≈	15.52	≈	15.52	
BOS11–16	158.81	≈	153.59	≈	150.12	150.95	148.98	≈	148.98	≈	148.98	≈	157.48	
BOS11–17	205.77	≈	197.54	≈	199.15	196.25	211.59	≈	211.59	≈	211.59	≈	228.36	
BOS11–26	309.39	≈	304.57	≈	304.15	299.85	327.16	≈	327.16	≈	327.16	<	374.79	
BOS11–27	227.82	≈	231.03	≈	234.28	235.88	224.85	≈	224.85	≈	224.85	≈	281.24	
BOS11–30	138.28	>	133.27	≈	131.54	129.86	139.46	≈	139.46	≈	139.46	≈	146.49	
BOS11–31	157.58	≈	156.52	≈	155.81	154.02	151.87	≈	151.87	≈	151.87	≈	196.99	
BK12–28	536.19	≈	532.08	≈	529.47	536.77	538.17	≈	538.17	≈	538.17	≈	567.99	
D05–18	350.65	>	338.4	≈	337.77	334.18	340.33	≈	340.33	≈	340.33	≈	350.59	
D05–19	366.67	≈	366.81	≈	364.88	361.33	380.66	≈	386.44	≈	388.49	≈	388.49	
DF11–6	532.38	≈	524.97	≈	526.02	526.15	579.84	≈	579.84	≈	579.84	<	747.77	
DF11–7	1316.59	≈	1310.02	≈	1309.99	1316.79	1359.89	≈	1359.89	≈	1359.89	≪	1566.58	
DF11–8	1092.7	≈	1082.36	≈	1082.77	1078.24	1028.93	≈	1028.93	≈	1028.93	<	1153.72	
DF11–22	928.2	≈	926.45	≈	928.99	930.3	1012.26	≈	1012.26	≈	1012.26	≈	1152.14	
DF11–23	776.48	≈	771.57	≈	770.87	767.04	723.5	≈	723.5	≈	723.5	≈	782.51	
DF11–24	479.3	≈	479.38	≈	471.25	483.25	450.61	≈	450.61	≈	450.61	<	530.97	
DF15–4	329.82	≈	322.21	≈	323.22	320.02	301.69	≈	301.69	≈	301.69	<	342.05	
DF15–5	1610.1	≈	1602.79	≈	1599.28	1606.96	1581.28	≈	1581.28	≈	1581.28	<	1712.9	
DF15–20	1222.77	≈	1231.06	≈	1229.59	1232.33	1273.14	≈	1273.14	≈	1273.14	≪	1582.66	
DF15–21	1575.91	≈	1587.09	≈	1571.12	1591.27	1664.01	≈	1664.01	≈	1664.01	<	1754.9	
DF15–33	2378.58	≈	2399.8	≈	2401.15	2405.23	2582.61	≈	2582.61	≈	2582.61	≪	2935.81	
DF15–35	642.04	≈	639.91	≈	637.62	641.02	722.6	≈	733.28	≈	733.28	≈	789.09	
DRFN08–10	232.84	≈	233.55	≈	223.78	244.91	276.61	≈	276.61	≈	276.61	≈	301.08	
DRFN08–11	354.52	≈	341.48	≈	340.95	341.42	336.45	≈	336.84	≈	339.09	≈	345.37	
DO09–32	1706.1	≈	1753.41	≈	1719.86	1761.6	1857.06	≈	1876.72	≈	1883.52	≈	2003.41	
FY17–25	422.32	≈	424.41	≈	419.44	423.34	433.18	≈	433.18	≈	433.18	<	464.23	
FRD12–29	452.64	≈	450.08	≈	447.25	452.6	484.5	≈	477.91	≈	514.87	≈	534.47	
KS13–12	1782.43	≈	1777.83	≈	1773.55	1775.62	1751.81	≈	1751.81	≈	1751.81	≈	1830.26	
STS13–13	959.21	≈	952.56	≈	949.46	951.34	955.73	≈	955.73	≈	955.73	≈	1023.43	
Aoyagi and Frechette (2009)	415.47	≈	421.18	>	409.19	423.68	416.51	≈	416.51	≈	416.51	≪	540.47	
Blonski et al. (2011)	1518.54	≫	1395.94	≈	1393.41	1346.79	1398.5	≈	1398.5	≈	1398.5	<	1564.48	
Bruttel and Kamecke (2012)	536.19	≈	532.08	≈	529.47	536.77	538.17	≈	538.17	≈	538.17	≈	567.99	
Dal Bó (2005)	727.25	≈	710.88	≈	708.32	699.05	726.04	≈	731.81	≈	732.27	≈	741.2	
Dal Bó and Fréchette (2011)	5201.05	≈	5137.82	≈	5132.96	5128.69	5195.88	≈	5195.88	≈	5195.88	≪	5960.78	
Dal Bó and Fréchette (2015)	7840.87	≈	7829.51	≈	7808.63	7825.98	8172.63	≈	8177.46	≈	8177.46	≪	9143.98	
Dreber et al. (2008)	597.17	≈	580.63	≈	570.33	589.84	618.5	≈	618.89	≈	619.9	≈	648.55	
Duffy and Ochs (2009)	1706.1	≈	1753.41	≈	1719.86	1761.6	1857.06	≈	1876.72	≈	1883.52	≈	2003.41	
Fréchette and Yuksel (2017)	422.32	≈	424.41	≈	419.44	423.34	433.18	≈	433.18	≈	433.18	<	464.23	
Fudenberg et al. (2012)	452.64	≈	450.08	≈	447.25	452.6	484.5	≈	477.91	≈	514.87	≈	534.47	
Kagel and Schley (2013)	1782.43	≈	1777.83	≈	1773.55	1775.62	1751.81	≈	1751.81	≈	1751.81	≈	1830.26	
Sherstyuk et al. (2013)	959.21	≈	952.56	≈	949.46	951.34	955.73	≈	955.73	≈	955.73	≈	1023.43	
Pooled	**22669.91**	≫	**22258.14**	≈	**22153.69**	**22097.67**	**22811.34**	≈	**22828.13**	≈	**22848.49**	≪	**25177.57**	

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 14.

Table 31: **Continuation strategies: Memory-1 or Memory-2, and semi-grim, pure or generalized pure?** Strategy mixtures are estimated treatment-by-treatment. The resulting ICL-BICs are pooled within experiments and overall (less is better, relation signs point to better models)

	Memory-2 Generalizations of Semi-Grim + AD						AD+SG		Best Mixtures of Generalized Pure Strategies						Best Pure
	M2"General"		M2"TFT"		M2"Grim"				M1+M2"TFT"		M1+M2"Grim"		M1		M1 & M2
Specification															
# Models evaluated	1		1		1		1		22^{32}		22^{32}		13^{32}		5^{32}
# Pars estimated (by treatment)	12		6		6		5		160		160		80		32
# Parameters accounted for	12		6		6		5		6–15		6–15		6–10		3–8
First halves per session															
Aoyagi and Frechette (2009)	692.5	$\approx$	690.85	$\approx$	686.2	$\approx$	694.72	$>$	649.38	$\approx$	646.7	$\approx$	645.31	$\ll$	791.38
Blonski et al. (2011)	714	$\gg$	601.67	$\approx$	601.95	$\gg$	549.45	$\ll$	760.28	$\approx$	768.65	$\gg$	713.8	$\approx$	703.06
Bruttel and Kamecke (2012)	572.14	$\approx$	566.75	$\approx$	567.58	$\approx$	567.86	$\approx$	569.25	$\approx$	576.94	$\approx$	585.42	$\approx$	588.54
Dal Bó (2005)	385.61	$>$	367.94	$\approx$	366.48	$\approx$	358.51	$\ll$	404.24	$\approx$	402.79	$\approx$	407.86	$\approx$	389.05
Dal Bó and Fréchette (2011)	3596.64	$\approx$	3542.28	$\approx$	3538.64	$\approx$	3533.99	$\approx$	3481.93	$\approx$	3517.82	$\approx$	3536.73	$\ll$	3835.73
Dal Bó and Fréchette (2015)	5017.27	$\approx$	4974.8	$\approx$	4988.94	$\approx$	4991.74	$\ll$	5193.81	$\approx$	5218.03	$\approx$	5259.64	$\ll$	5538.35
Dreber et al. (2008)	464.84	$>$	444.11	$\approx$	444.71	$\approx$	437.17	$\approx$	474.89	$\approx$	483.06	$\approx$	478.09	$\approx$	462.72
Duffy and Ochs (2009)	1060.26	$\approx$	1063.66	$\approx$	1074.9	$\approx$	1090.22	$\approx$	1039.24	$\approx$	1045.68	$\approx$	1047.59	$\approx$	1102.63
Fréchette and Yuksel (2017)	174.64	$\approx$	167.06	$\approx$	164.75	$\approx$	161.45	$\ll$	182.7	$\approx$	188.04	$\approx$	188.5	$\approx$	181.98
Fudenberg et al. (2012)	301.76	$\approx$	293.52	$\approx$	294.4	$\approx$	291.43	$<$	319.76	$\approx$	322.89	$\approx$	319.45	$<$	366.77
Kagel and Schley (2013)	1746.26	$\approx$	1749.95	$\approx$	1753.68	$\approx$	1782.82	$>$	1651.6	$\approx$	1679.19	$<$	1761.98	$\approx$	1805.96
Sherstyuk et al. (2013)	917.07	$\approx$	907.95	$\approx$	913.52	$\approx$	912.8	$\approx$	857.56	$\approx$	870.11	$\approx$	865.67	$<$	941.92
Pooled	**16080.69**	$\gg$	**15589.39**	$\approx$	**15614.59**	$>$	**15481.59**	$\ll$	**15948.02**	$<$	**16109.87**	$\approx$	**16077.95**	$\ll$	**16858.76**
Second halves per session															
Aoyagi and Frechette (2009)	396.32	$\approx$	391.42	$>$	387.48	$\approx$	389.24	$\approx$	368.08	$\approx$	365.6	$\approx$	363.58	$\ll$	484.41
Blonski et al. (2011)	1012.48	$\gg$	919.29	$\approx$	922.48	$\gg$	867.87	$<$	1007.71	$\approx$	1021.03	$>$	992.44	$\approx$	1055.98
Bruttel and Kamecke (2012)	333.51	$\approx$	337.12	$\approx$	329.73	$\approx$	347.4	$\approx$	318.55	$\approx$	328.38	$\approx$	344.88	$\approx$	316.37
Dal Bó (2005)	449.03	$\approx$	434.38	$\approx$	433.82	$\approx$	424.44	$<$	451.25	$<$	471.53	$\approx$	475.11	$\approx$	463.53
Dal Bó and Fréchette (2011)	2854.52	$\approx$	2801.46	$\approx$	2800.71	$\approx$	2817.31	$>$	2619.87	$\approx$	2643.97	$<$	2737.11	$<$	2885.4
Dal Bó and Fréchette (2015)	5006.3	$\approx$	5013.49	$\approx$	5012.99	$\approx$	5043.81	$\approx$	5034.8	$\approx$	5099.68	$\approx$	5164.78	$\ll$	5575.88
Dreber et al. (2008)	272.94	$\approx$	258.88	$\approx$	253.47	$\approx$	264.94	$\approx$	287.55	$\approx$	288.95	$\approx$	295.06	$\approx$	287.58
Duffy and Ochs (2009)	1375.43	$\approx$	1367.68	$\approx$	1389.92	$\approx$	1403.03	$\approx$	1339.22	$\approx$	1359.29	$\approx$	1381.01	$\ll$	1617.76
Fréchette and Yuksel (2017)	308.21	$\approx$	304.2	$\approx$	306.93	$\approx$	313.5	$\approx$	311.03	$\approx$	311.99	$\approx$	309.63	$\ll$	356.11
Fudenberg et al. (2012)	384.37	$\approx$	382.32	$\approx$	378.59	$\approx$	380.75	$\approx$	367.2	$\approx$	364.46	$\approx$	373.44	$<$	447.18
Kagel and Schley (2013)	1204.38	$\approx$	1202.61	$\approx$	1197.19	$\approx$	1211.37	$>$	1088.27	$\approx$	1122.98	$\approx$	1170.12	$\approx$	1169.32
Sherstyuk et al. (2013)	598.79	$\approx$	590.65	$\approx$	591.38	$\approx$	586.72	$>$	503.98	$\approx$	517.77	$\approx$	527.09	$\approx$	583.8
Pooled	**14633.97**	$\gg$	**14222.35**	$\approx$	**14223.53**	$\approx$	**14159.8**	$\approx$	**14059.75**	$<$	**14267.75**	$\approx$	**14387.48**	$\ll$	**15400.68**

Note: Results treatment-by-treatment are in the appendix. The main body contains ICL-BICs aggregated at paper level. Relation signs and p-values are exactly as above, see Table 3. "M2" ("M1") denotes strategies, whose actions may depend on actions in $t-2$ and $t-1$ ($t-1$ only). The supplements "General", "TFT", "Grim" indicate whether parameters of behavior strategies may depend on: all four possible histories in $t-2$ (M2 "General"), whether the opponent cooperated in $t-2$ (M2 "TFT"), or whether there was joint cooperation in $t-2$ (M2 "Grim"). Pure M2 strategies do not have such free parameters. Columns 1-3 contain one memory-2 version of semi-grim each. Column 4 is memory-1 semi-grim. Columns 5-7 are memory-2 and memory-1 versions of generalized prototypical strategies. The last column contains the best fitting combinations of a set of pure memory-1 and memory-2 strategies from the literature (TFT, Grim, AD, Grim2, TF2T, T2, 2TFT, 2PTFT) for definitions see Table 12 in the Online Appendix.

Table 32: Is there a single "semi grim" type? Mixture models involving SG

	Best Mixture Best Switching		SG + AD		1.5×SG+AD		2×SG+AD		3×SG+AD		3×P5+AD		2×P5+AD		P5+AD
Specification															
# Models evaluated	$39^{32} \approx 10^{51}$		1		1		1		1		1		1		1
# Pars estimated (by treatment)	438		5		7		9		13		19		17		11
# Parameters accounted for	3-30		5		7		9		13		19		17		11
First halves per session															
Aoyagi and Frechette (2009)	755.97	$\approx$	781.86	$\approx$	792.51	$\approx$	777.3	$\approx$	782.13	$>$	741.95	$\approx$	744.86	$\approx$	744.06
Blonski et al. (2011)	1069.39	$\approx$	1069.28	$\approx$	1104.6	$\approx$	1134.96	$\ll$	1232.97	$\ll$	1332.48	$\gg$	1205.47	$\gg$	1106.01
Bruttel and Kamecke (2012)	785.49	$\approx$	800.12	$\approx$	771.14	$\approx$	762.83	$\approx$	748.06	$\approx$	751.86	$\approx$	759.45	$\approx$	803.58
Dal Bó (2005)	631.2	$\approx$	629.17	$\approx$	618.39	$\approx$	600.26	$\ll$	626.56	$\approx$	639.8	$>$	609.1	$\approx$	620.38
Dal Bó and Fréchette (2011)	6388.49	$<$	6597.93	$>$	6352.59	$\approx$	6304.97	$\approx$	6198.12	$\approx$	6216.22	$<$	6295.32	$\ll$	6553.25
Dal Bó and Fréchette (2015)	8138.61	$\approx$	8017.59	$\gg$	7830.12	$\approx$	7810.7	$\approx$	7828.38	$\approx$	7829.74	$\approx$	7775.7	$\ll$	7969.32
Dreber et al. (2008)	752.16	$\approx$	782.37	$\approx$	764.44	$\approx$	763.52	$\approx$	766.77	$\approx$	765.81	$\approx$	767.32	$\approx$	783.45
Duffy and Ochs (2009)	1372.99	$\approx$	1372.97	$\approx$	1361.15	$\approx$	1320.71	$\approx$	1297.84	$\approx$	1291.42	$<$	1345.16	$\approx$	1361.86
Fréchette and Yuksel (2017)	298.53	$\approx$	299.62	$\approx$	289.54	$\approx$	284.11	$\approx$	289.88	$\approx$	294.05	$\approx$	285.33	$\approx$	291.69
Fudenberg et al. (2012)	425.54	$>$	381.01	$\approx$	377.96	$\approx$	370.01	$\approx$	380.86	$\approx$	381.34	$\approx$	372.32	$\approx$	377.33
Kagel and Schley (2013)	2439.06	$\approx$	2561.76	$\gg$	2450.24	$\approx$	2421.27	$\approx$	2385.02	$\approx$	2354.05	$\approx$	2398.74	$\ll$	2551.68
Sherstyuk et al. (2013)	1296.85	$\approx$	1303.8	$\approx$	1234.52	$\approx$	1200.28	$\approx$	1184.82	$\approx$	1177.24	$\approx$	1186.92	$\ll$	1286.14
Pooled	**24863.15**	$\approx$	**24779.85**	$\gg$	**24202.51**	$\approx$	**24079.18**	$\approx$	**24195.57**	$<$	**24468.99**	$>$	**24219.87**	$\ll$	**24704.09**
Second halves per session															
Aoyagi and Frechette (2009)	416.51	$\approx$	423.68	$\approx$	421.21	$\approx$	422.24	$\approx$	423.63	$>$	404.95	$\approx$	408.59	$\approx$	409.04
Blonski et al. (2011)	1394.16	$\approx$	1346.79	$\approx$	1370.16	$\approx$	1385.91	$<$	1442.85	$\ll$	1555.48	$\gg$	1453.1	$\gg$	1379.87
Bruttel and Kamecke (2012)	516.71	$\approx$	536.77	$\gg$	480.47	$\approx$	478.23	$\approx$	470.25	$\approx$	443.83	$\approx$	471.73	$<$	528.54
Dal Bó (2005)	729.48	$>$	699.05	$\approx$	677.24	$\approx$	679.04	$<$	697.21	$\approx$	707.25	$\approx$	687.86	$\approx$	696.41
Dal Bó and Fréchette (2011)	4964.77	$\approx$	5128.69	$\gg$	4565.93	$\approx$	4545.08	$\approx$	4426.48	$\approx$	4461.98	$\approx$	4493.1	$\ll$	5045.34
Dal Bó and Fréchette (2015)	7893.79	$\approx$	7825.98	$\gg$	7306.25	$\approx$	7310.27	$>$	7170.25	$\approx$	7089.56	$\approx$	7151.84	$\ll$	7683.76
Dreber et al. (2008)	581.94	$\approx$	589.84	$>$	544.66	$\approx$	541.83	$\approx$	539.47	$\approx$	519.28	$\approx$	518.82	$<$	562.99
Duffy and Ochs (2009)	1850.35	$\approx$	1761.6	$\gg$	1656.55	$\approx$	1602.93	$>$	1518.65	$\approx$	1509.7	$<$	1598.04	$\ll$	1715.88
Fréchette and Yuksel (2017)	427.79	$\approx$	423.34	$\approx$	422.61	$\approx$	381.63	$\approx$	375.03	$\approx$	384.11	$\approx$	382.16	$<$	409.93
Fudenberg et al. (2012)	514.87	$\approx$	452.6	$\approx$	433.74	$\approx$	414.24	$\approx$	405.22	$\approx$	410.69	$\approx$	421.81	$\approx$	448.37
Kagel and Schley (2013)	1678.7	$\approx$	1775.62	$\gg$	1572.95	$\approx$	1541.38	$>$	1488.49	$\approx$	1477.87	$\approx$	1527.47	$\ll$	1748.01
Sherstyuk et al. (2013)	955.73	$\approx$	951.34	$\gg$	834.74	$\approx$	823.06	$\approx$	799.39	$\approx$	801.53	$\approx$	815.26	$\ll$	935.01
Pooled	**22422.07**	$\approx$	**22097.67**	$\gg$	**20541.83**	$\approx$	**20454.13**	$>$	**20231.09**	$<$	**20459.26**	$\approx$	**20403.95**	$\ll$	**21818.45**

Note: This table verifies a number of possible mixtures involving Semi-Grim types as a robustness check for the sufficiency of focussing on the mixtures examined above. E.g. "3× SG refers to a model containing three different versions of memory-1 semi-grim with allowing for heterogeneity of randomization parameters across subjects.

Table 33: Table 32 by treatments – Is there a single "semi grim" type? Mixture models involving SG

(a) First halves per session

	Best Mixture Best Switching		SG + AD		1.5×SG+AD		2×SG+AD		3×SG+AD		3×P5+AD		2×P5+AD		P5+AD
Specification															
# Models evaluated	$39^{32} \approx 10^{51}$		1		1		1		1		1		1		1
# Pars estimated (by treatment)	438		5		7		9		13		19		17		11
# Parameters accounted for	3-30		5		7		9		13		19		17		11
AF09–34	755.97	≈	781.86	≈	792.51	≈	777.3	≈	782.13	>	741.95	≈	744.86	≈	744.06
BOS11–9	83.42	≈	86.56	≈	88.35	≈	85.71	<	92	<	97.71	≈	91.58	≈	89.57
BOS11–14	90	≈	93.88	≈	98.01	≈	95.87	≪	105.42	<	113.81	≫	99.92	≈	96.86
BOS11–15	32.69	≈	37.73	<	43.07	≪	53.61	≪	61.19	≪	72.34	≫	60.83	≫	42.63
BOS11–16	167.3	≈	167.42	≈	157.13	≈	157.72	≈	162.97	≈	171.76	≈	165.4	≈	169.7
BOS11–17	110.57	≈	115.02	≈	119.79	≈	122.41	≈	129.42	≈	123.86	≈	124.68	>	112.95
BOS11–26	256.37	≈	244.5	≈	246.46	≈	249.78	≈	248.26	≈	247.15	≈	245.48	≈	245.83
BOS11–27	102.11	≈	92.83	≈	92.07	≈	93.42	<	103.66	≈	106.06	≫	93.17	≈	91.86
BOS11–30	56.81	≈	55.74	<	61.12	≈	64.33	<	73.55	≪	82.55	≫	70.18	≫	58.74
BOS11–31	125.82	≈	125.52	≈	128.49	≈	121.98	≈	126.3	≈	126.95	≈	124.03	≈	127.75
BK12–28	785.49	≈	800.12	≈	771.14	≈	762.83	≈	748.06	≈	751.86	≈	759.45	≈	803.58
D05–18	230.66	≈	230.57	≈	238.35	≈	229.01	<	241.65	≈	249.85	≈	238.41	≈	233.85
D05–19	396.28	≈	395.06	≈	375.07	≈	364.87	≈	375.69	≈	376.48	≈	361.48	<	381.57
DF11–6	770.36	≈	794.44	≈	748.9	≈	752.56	≈	730.86	≈	734.21	≈	753.19	≈	793.66
DF11–7	1132.04	<	1256.83	≈	1229.61	≈	1238	≈	1219.19	≈	1214.51	≈	1243.2	≈	1256.83
DF11–8	1279.8	<	1389.06	≈	1286.68	≈	1289.21	≈	1255.59	≈	1249.19	<	1279.81	≪	1379.72
DF11–22	1056.77	>	965.96	≈	961.58	≈	944.2	≈	934.1	≈	934.68	≈	945.32	≈	963.43
DF11–23	1020.09	≈	1019.57	≈	972.17	≈	941.73	≈	921.62	≈	919.31	≈	937.03	≈	1003.42
DF11–24	1022.62	≪	1145.15	≈	1115.95	≈	1090.82	≈	1066.75	≈	1062.01	≈	1066.78	≈	1118.5
DF15–4	395.89	≈	436.57	≈	422.43	≈	425	≈	433.05	≈	423.74	≈	421.84	≈	426.78
DF15–5	1638.92	≪	1738.77	≫	1649.1	≈	1639.52	≈	1632.23	≈	1637.48	≈	1637.39	≪	1735.05
DF15–20	1433.87	≈	1405.3	>	1366.09	≈	1364.37	≈	1365.64	≈	1341.25	≈	1352.57	≈	1392.99
DF15–21	1902.95	≈	1872.47	>	1827.37	≈	1811.11	≈	1806.23	≈	1811.27	≈	1792.58	≈	1852.64
DF15–33	2296.41	>	2186.26	≈	2178.12	≈	2174.56	≈	2165.14	≈	2158.49	≈	2161.68	≈	2179.74
DF15–35	372.62	≈	349.06	≈	346.18	≈	343.65	≈	350.27	≈	346.71	≈	333.83	≈	341.3
DRFN08–10	334.73	<	367.86	≈	359.04	≈	358.63	≈	358.23	≈	357.95	≈	359.19	≈	367.27
DRFN08–11	405.73	≈	411.01	≈	400.49	≈	398.59	≈	399.43	≈	394.56	≈	399.03	≈	411.28
DO09–32	1372.99	≈	1372.97	≈	1361.15	≈	1320.71	≈	1297.84	≈	1291.42	<	1345.16	≈	1361.86
FY17–25	298.53	≈	299.62	≈	289.54	≈	284.11	≈	289.88	≈	294.05	≈	285.33	≈	291.69
FRD12–29	425.54	>	381.01	≈	377.96	≈	370.01	≈	380.86	≈	381.34	≈	372.32	≈	377.33
KS13–12	2439.06	≈	2561.76	≫	2450.24	≈	2421.27	≈	2385.02	≈	2354.05	≈	2398.74	≪	2551.68
STS13–13	1296.85	≈	1303.8	≈	1234.52	≈	1200.28	≈	1184.82	≈	1177.24	≈	1186.92	≪	1286.14
Aoyagi and Frechette (2009)	755.97	≈	781.86	≈	792.51	≈	777.3	≈	782.13	>	741.95	≈	744.86	≈	744.06
Blonski et al. (2011)	1069.39	≈	1069.28	≈	1104.6	≈	1134.96	≪	1232.97	≪	1332.48	≫	1205.47	≫	1106.01
Bruttel and Kamecke (2012)	785.49	≈	800.12	≈	771.14	≈	762.83	≈	748.06	≈	751.86	≈	759.45	≈	803.58
Dal Bó (2005)	631.2	≈	629.17	≈	618.39	≈	600.26	≪	626.56	≈	639.8	>	609.1	≈	620.38
Dal Bó and Fréchette (2011)	6388.49	<	6597.93	>	6352.59	≈	6304.97	≈	6198.12	≈	6216.22	<	6295.32	≪	6553.25
Dal Bó and Fréchette (2015)	8138.61	≈	8017.59	≫	7830.12	≈	7810.7	≈	7828.38	≈	7829.74	≈	7775.7	≪	7969.32
Dreber et al. (2008)	752.16	≈	782.37	≈	764.44	≈	763.52	≈	766.77	≈	765.81	≈	767.32	≈	783.45
Duffy and Ochs (2009)	1372.99	≈	1372.97	≈	1361.15	≈	1320.71	≈	1297.84	≈	1291.42	<	1345.16	≈	1361.86
Fréchette and Yuksel (2017)	298.53	≈	299.62	≈	289.54	≈	284.11	≈	289.88	≈	294.05	≈	285.33	≈	291.69
Fudenberg et al. (2012)	425.54	>	381.01	≈	377.96	≈	370.01	≈	380.86	≈	381.34	≈	372.32	≈	377.33
Kagel and Schley (2013)	2439.06	≈	2561.76	≫	2450.24	≈	2421.27	≈	2385.02	≈	2354.05	≈	2398.74	≪	2551.68
Sherstyuk et al. (2013)	1296.85	≈	1303.8	≈	1234.52	≈	1200.28	≈	1184.82	≈	1177.24	≈	1186.92	≪	1286.14
Pooled	**24863.15**	≈	**24779.85**	≫	**24202.51**	≈	**24079.18**	≈	**24195.57**	<	**24468.99**	>	**24219.87**	≪	**24704.09**

(b) Second halves per session

	Best Mixture Best Switching		SG + AD		1.5×SG+AD		2×SG+AD		3×SG+AD		3×P5+AD		2×P5+AD		P5+AD
Specification															
# Models evaluated	$39^{32} \approx 10^{51}$		1		1		1		1		1		1		1
# Pars estimated (by treatment)	438		5		7		9		13		19		17		11
# Parameters accounted for	3-30		5		7		9		13		19		17		11
AF09–34	416.51	≈	423.68	≈	421.21	≈	422.24	≈	423.63	>	404.95	≈	408.59	≈	409.04
BOS11–9	78.84	≈	75.1	≈	80.12	≈	76.06	<	81.99	≈	86.75	>	77.82	≈	78.45
BOS11–14	40.82	≈	35.58	≈	35.86	≪	38.21	≪	44.97	≪	61.2	≫	51.58	≈	43.6
BOS11–15	15.52	≈	19.23	<	24.71	≪	30.15	≪	37.6	≪	53.63	≫	36.89	≫	18.77
BOS11–16	148.98	≈	150.95	≈	138.89	≈	149.71	≈	144.32	≈	145.7	≈	143.01	≈	151.37
BOS11–17	211.59	≈	196.25	≈	201.03	≈	198.59	≈	205.53	≈	205.55	≈	199.44	≈	196.67
BOS11–26	327.16	≈	299.85	≈	309.63	≈	295.25	≈	301.95	≈	305.72	≈	300.57	≈	301.47
BOS11–27	224.85	≈	235.88	≈	223.91	≈	224.21	≈	212.63	≈	212.04	≈	222.13	≈	234.53
BOS11–30	137.43	≈	129.86	≈	132.45	≈	133.81	≈	137.88	≈	146.11	≫	134.06	≈	130.39
BOS11–31	151.87	≈	154.02	≈	153.45	≈	149.77	≈	145.78	≈	148.5	≈	157.41	≈	154.5
BK12–28	516.71	≈	536.77	≫	480.47	≈	478.23	≈	470.25	≈	443.83	≈	471.73	<	528.54
D05–18	340.33	≈	334.18	>	312.74	≈	315.38	≈	323.23	≈	323.92	≈	314.03	≈	330.79
D05–19	383.3	≈	361.33	≈	359.54	≈	357.27	≈	364.76	≈	369.86	≈	364.62	≈	360.66
DF11–6	557.16	≈	526.15	≈	489.74	≈	509.84	≈	495.28	≈	492.63	≈	484.7	≈	516.27
DF11–7	1268.34	≈	1316.79	≈	1250.02	≈	1246.93	≈	1197.33	≈	1212.7	≈	1235.02	≈	1305.98
DF11–8	904.89	≪	1078.24	>	871.84	≈	872.45	≈	834.02	<	852.42	<	869.39	≪	1065.6
DF11–22	973.62	≈	930.3	>	858.03	≈	860.39	≈	848.36	≈	832.44	≈	844.82	<	918.07
DF11–23	723.5	≈	767.04	>	608.87	≈	582.24	≈	556.55	≈	545.64	≈	564.18	<	741.18
DF11–24	450.61	≈	483.25	≈	449.72	≈	424.76	≈	424.94	≈	423.85	≈	424.99	≈	460.55
DF15–4	301.69	≈	320.02	≈	299.49	≈	306.89	≈	301.34	≈	299.43	≈	295.28	≈	318.1
DF15–5	1435.63	≪	1606.96	≫	1407.26	≈	1409.52	≈	1395.3	≈	1409.68	≈	1407.43	≪	1596.73
DF15–20	1270.1	≈	1232.33	≫	1145.96	≈	1141.29	≈	1113.17	≈	1107.65	≈	1121.22	<	1215.9
DF15–21	1504.63	≈	1591.27	≫	1453.74	≈	1422.87	≈	1390.81	≈	1367.29	≈	1403.53	≪	1559.69
DF15–33	2541.5	>	2405.23	≈	2331.38	≈	2349.48	>	2265.94	>	2184.74	<	2237.97	<	2318.99
DF15–35	723.78	>	641.02	≈	627.6	≈	627.73	≈	627.87	≈	609.97	≈	610.6	≈	633.52
DRFN08–10	243.71	≈	244.91	≈	234.76	≈	236.44	≈	234.24	≈	218.82	≈	218.68	≈	225.25
DRFN08–11	323.06	≈	341.42	>	305	≈	299.09	≈	296.13	≈	287.15	≈	291.03	≈	332.84
DO09–32	1850.35	≈	1761.6	≫	1656.55	≈	1602.93	>	1518.65	≈	1509.7	<	1598.04	≪	1715.88
FY17–25	427.79	≈	423.34	≈	422.61	≈	381.63	≈	375.03	≈	384.11	≈	382.16	<	409.93
FRD12–29	514.87	≈	452.6	≈	433.74	≈	414.24	≈	405.22	≈	410.69	≈	421.81	≈	448.37
KS13–12	1678.7	≈	1775.62	≫	1572.95	≈	1541.38	>	1488.49	≈	1477.87	≈	1527.47	≪	1748.01
STS13–13	955.73	≈	951.34	≫	834.74	≈	823.06	≈	799.39	≈	801.53	≈	815.26	≪	935.01
Aoyagi and Frechette (2009)	416.51	≈	423.68	≈	421.21	≈	422.24	≈	423.63	>	404.95	≈	408.59	≈	409.04
Blonski et al. (2011)	1394.16	≈	1346.79	≈	1370.16	≈	1385.91	<	1442.85	≪	1555.48	≫	1453.1	≫	1379.87
Bruttel and Kamecke (2012)	516.71	≈	536.77	≫	480.47	≈	478.23	≈	470.25	≈	443.83	≈	471.73	<	528.54
Dal Bó (2005)	729.48	>	699.05	≈	677.24	≈	679.04	<	697.21	≈	707.25	≈	687.86	≈	696.41
Dal Bó and Fréchette (2011)	4964.77	≈	5128.69	≫	4565.93	≈	4545.08	≈	4426.48	≈	4461.98	≈	4493.1	≪	5045.34
Dal Bó and Fréchette (2015)	7893.79	≈	7825.98	≫	7306.25	≈	7310.27	>	7170.25	≈	7089.56	≈	7151.84	≪	7683.76
Dreber et al. (2008)	581.94	≈	589.84	>	544.66	≈	541.83	≈	539.47	≈	519.28	≈	518.82	<	562.99
Duffy and Ochs (2009)	1850.35	≈	1761.6	≫	1656.55	≈	1602.93	>	1518.65	≈	1509.7	<	1598.04	≪	1715.88
Fréchette and Yuksel (2017)	427.79	≈	423.34	≈	422.61	≈	381.63	≈	375.03	≈	384.11	≈	382.16	<	409.93
Fudenberg et al. (2012)	514.87	≈	452.6	≈	433.74	≈	414.24	≈	405.22	≈	410.69	≈	421.81	≈	448.37
Kagel and Schley (2013)	1678.7	≈	1775.62	≫	1572.95	≈	1541.38	>	1488.49	≈	1477.87	≈	1527.47	≪	1748.01
Sherstyuk et al. (2013)	955.73	≈	951.34	≫	834.74	≈	823.06	≈	799.39	≈	801.53	≈	815.26	≪	935.01
Pooled	**22422.07**	≈	**22097.67**	≫	**20541.83**	≈	**20454.13**	>	**20231.09**	<	**20459.26**	≈	**20403.95**	≪	**21818.45**

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 14.

Table 34: Strategies as a function of behavior in $t-2$ (TFT scheme)

Experiment	Cooperation after $\emptyset, (c,c), (d,c)$ in $t-2$								Cooperation after $(c,d), (d,d)$ in $t-2$							
	$\hat{\sigma}_{cc}$		$\hat{\sigma}_{cd}$		$\hat{\sigma}_{dc}$		$\hat{\sigma}_{dd}$		$\hat{\sigma}_{cc}$		$\hat{\sigma}_{cd}$		$\hat{\sigma}_{dc}$		$\hat{\sigma}_{dd}$	
First halves per session																
Aoyagi and Frechette (2009)	0.93	$\gg$	0.439	$\approx$	0.388	$\approx$	0.434		0.789	$\gg$	0.463	$\approx$	0.44	$>$	0.291	
Blonski et al. (2011)	0.901	$\gg$	0.27	$\approx$	0.146	$\gg$	0.053		0.667	$\approx$	0.296	$\approx$	0.321	$\gg$	0.027	
Bruttel and Kamecke (2012)	0.908	$\gg$	0.312	$\approx$	0.218	$\approx$	0.151		0.944	$\gg$	0.247	$\approx$	0.247	$\gg$	0.063	
Dal Bó (2005)	0.93	$\gg$	0.232	$\approx$	0.31	$>$	0.126		0.833	$>$	0.147	$\approx$	0.413	$\gg$	0.071	
Dal Bó and Fréchette (2011)	0.955	$\gg$	0.352	$\approx$	0.298	$\gg$	0.086		0.885	$\gg$	0.291	$\approx$	0.41	$\gg$	0.048	
Dal Bó and Fréchette (2015)	0.944	$\gg$	0.301	$\approx$	0.277	$\gg$	0.098		0.847	$\gg$	0.288	$\approx$	0.44	$\gg$	0.044	
Dreber et al. (2008)	0.902	$\gg$	0.213	$\approx$	0.189	$\gg$	0.061		1	$>$	0.233	$\approx$	0.302	$\gg$	0.025	
Duffy and Ochs (2009)	0.927	$\gg$	0.316	$\approx$	0.304	$\approx$	0.232		0.691	$\gg$	0.277	$\approx$	0.361	$\gg$	0.08	
Fréchette and Yuksel (2017)	0.943	$\gg$	0.153	$\approx$	0.241	$\approx$	0.1		1	$\approx$		$\approx$	0.4	$\approx$	0.086	
Fudenberg et al. (2012)	0.984	$\gg$	0.394	$\approx$	0.347	$\gg$	0.05		0.895	$\gg$	0.41	$\approx$	0.579	$\gg$	0.069	
Kagel and Schley (2013)	0.94	$\gg$	0.29	$\approx$	0.25	$\gg$	0.125		0.787	$\gg$	0.196	$\approx$	0.402	$\gg$	0.032	
Sherstyuk et al. (2013)	0.951	$\gg$	0.329	$\approx$	0.341	$>$	0.186		0.844	$\gg$	0.328	$\approx$	0.424	$\gg$	0.09	
Pooled	0.944	$\gg$	0.312	$>$	0.279	$\gg$	0.106		0.826	$\gg$	0.287	$\approx$	0.41	$\gg$	0.05	
Second halves per session																
Aoyagi and Frechette (2009)	0.961	$\gg$	0.408	$\approx$	0.567	$\approx$	0.447		0.867	$\gg$	0.381	$\approx$	0.451	$\approx$	0.328	
Blonski et al. (2011)	0.922	$\gg$	0.224	$\approx$	0.195	$\gg$	0.029		0.944	$\gg$	0.402	$\approx$	0.324	$\gg$	0.018	
Bruttel and Kamecke (2012)	0.948	$\gg$	0.239	$\approx$	0.214	$\approx$	0.118		0.923	$>$	0.167	$\approx$	0.5	$\gg$	0.018	
Dal Bó (2005)	0.919	$\gg$	0.264	$\approx$	0.39	$\gg$	0.113		0.938	$\gg$	0.175	$\approx$	0.383	$\gg$	0.047	
Dal Bó and Fréchette (2011)	0.979	$\gg$	0.391	$\approx$	0.29	$\gg$	0.075		0.975	$\gg$	0.334	$\approx$	0.547	$\gg$	0.022	
Dal Bó and Fréchette (2015)	0.977	$\gg$	0.304	$\approx$	0.328	$\gg$	0.064		0.927	$\gg$	0.343	$\approx$	0.532	$\gg$	0.028	
Dreber et al. (2008)	0.917	$\gg$	0.111	$<$	0.311	$\gg$	0.005		0.909	$>$	0.5	$\approx$	0.629	$\gg$	0.01	
Duffy and Ochs (2009)	0.98	$\gg$	0.408	$\approx$	0.371	$>$	0.232		0.849	$\gg$	0.316	$\approx$	0.415	$\gg$	0.058	
Fréchette and Yuksel (2017)	0.973	$\gg$	0.213	$\approx$	0.286	$\approx$	0.214		0.818	$\approx$	0.286	$\approx$	0.575	$\gg$	0.038	
Fudenberg et al. (2012)	0.974	$\gg$	0.5	$\approx$	0.41	$\gg$	0.111		0.84	$>$	0.463	$\approx$	0.417	$\gg$	0.075	
Kagel and Schley (2013)	0.967	$\gg$	0.281	$\approx$	0.263	$\gg$	0.061		0.872	$\gg$	0.188	$\approx$	0.527	$\gg$	0.018	
Sherstyuk et al. (2013)	0.973	$\gg$	0.503	$\approx$	0.417	$\gg$	0.12		0.968	$\gg$	0.431	$\approx$	0.5	$\gg$	0.062	
Pooled	0.973	$\gg$	0.325	$\approx$	0.315	$\gg$	0.076		0.917	$\gg$	0.332	$\approx$	0.499	$\gg$	0.028	

Note: Relation signs, bootstrap procedure, and derived p-values are exactly as above, see Table 2, with the obvious adaptation that the Holm-Bonferroni correction now applies to all eight tests per data set.

Table 35: Strategies as a function of behavior in $t-2$ (Grim scheme)

Experiment	Cooperation after $\emptyset, (c,c)$ in $t-2$								Cooperation after $(c,d), (d,c), (d,d)$ in $t-2$							
	$\hat{\sigma}_{cc}$		$\hat{\sigma}_{cd}$		$\hat{\sigma}_{dc}$		$\hat{\sigma}_{dd}$		$\hat{\sigma}_{cc}$		$\hat{\sigma}_{cd}$		$\hat{\sigma}_{dc}$		$\hat{\sigma}_{dd}$	
First halves per session																
Aoyagi and Frechette (2009)	0.939	$\gg$	0.39	$\approx$	0.439	$\approx$	0.556		0.782	$\gg$	0.485	$\approx$	0.39	$>$	0.32	
Blonski et al. (2011)	0.903	$\gg$	0.248	$\approx$	0.174	$\gg$	0.045		0.714	$>$	0.318	$\approx$	0.216	$>$	0.031	
Bruttel and Kamecke (2012)	0.919	$\gg$	0.296	$\approx$	0.245	$\approx$	0.179		0.833	$\gg$	0.278	$\approx$	0.213	$\gg$	0.071	
Dal Bó (2005)	0.926	$\gg$	0.184	$\approx$	0.31	$\approx$	0.143		0.889	$\gg$	0.254	$\approx$	0.39	$\gg$	0.074	
Dal Bó and Fréchette (2011)	0.961	$\gg$	0.342	$\approx$	0.307	$\gg$	0.081		0.849	$\gg$	0.324	$\approx$	0.364	$\gg$	0.054	
Dal Bó and Fréchette (2015)	0.95	$\gg$	0.265	$\approx$	0.301	$\gg$	0.081		0.843	$\gg$	0.328	$\approx$	0.369	$\gg$	0.052	
Dreber et al. (2008)	0.901	$\gg$	0.154	$\approx$	0.217	$\gg$	0.062		1	$\gg$	0.359	$\approx$	0.203	$\gg$	0.031	
Duffy and Ochs (2009)	0.932	$\gg$	0.218	$\approx$	0.301	$\approx$	0.208		0.748	$\gg$	0.361	$\approx$	0.35	$\gg$	0.102	
Fréchette and Yuksel (2017)	0.942	$\gg$	0.132	$\approx$	0.245	$\gg$	0		1	$\approx$	0.182	$\approx$	0.364	$\approx$	0.111	
Fudenberg et al. (2012)	0.985	$\gg$	0.429	$\approx$	0.408	$\gg$	0		0.921	$\gg$	0.377	$\approx$	0.443	$\gg$	0.068	
Kagel and Schley (2013)	0.947	$\gg$	0.236	$\approx$	0.288	$\gg$	0.133		0.763	$\gg$	0.298	$\approx$	0.305	$\gg$	0.042	
Sherstyuk et al. (2013)	0.953	$\gg$	0.312	$\approx$	0.395	$\gg$	0.172		0.875	$\gg$	0.343	$\approx$	0.349	$\gg$	0.107	
Pooled	0.949	$\gg$	0.278	$\approx$	0.3	$\gg$	0.091		0.825	$\gg$	0.333	$\approx$	0.346	$\gg$	0.059	
Second halves per session																
Aoyagi and Frechette (2009)	0.965	$\gg$	0.438	$\approx$	0.625	$\approx$	0.333		0.846	$\gg$	0.371	$<$	0.443	$\approx$	0.378	
Blonski et al. (2011)	0.922	$\gg$	0.157	$\approx$	0.232	$\gg$	0.027		0.941	$\gg$	0.425	$\approx$	0.23	$\gg$	0.019	
Bruttel and Kamecke (2012)	0.946	$\gg$	0.156	$\approx$	0.233	$\approx$	0.173		0.958	$\gg$	0.327	$\approx$	0.4	$\gg$	0.019	
Dal Bó (2005)	0.918	$\gg$	0.178	$<$	0.4	$>$	0.131		0.937	$\gg$	0.32	$\approx$	0.373	$\gg$	0.052	
Dal Bó and Fréchette (2011)	0.981	$\gg$	0.373	$\approx$	0.323	$\gg$	0.077		0.95	$\gg$	0.38	$\approx$	0.416	$\gg$	0.025	
Dal Bó and Fréchette (2015)	0.98	$\gg$	0.264	$<$	0.366	$\gg$	0.058		0.904	$\gg$	0.369	$\approx$	0.44	$\gg$	0.031	
Dreber et al. (2008)	0.913	$\gg$	0.029	$\ll$	0.314	$\gg$	0.007		0.955	$\gg$	0.417	$\approx$	0.611	$\gg$	0.009	
Duffy and Ochs (2009)	0.981	$\gg$	0.362	$\approx$	0.433	$\approx$	0.226		0.889	$\gg$	0.369	$\approx$	0.368	$\gg$	0.077	
Fréchette and Yuksel (2017)	0.976	$\gg$	0.173	$\approx$	0.308	$\approx$	0.222		0.75	$>$	0.294	$\approx$	0.49	$\gg$	0.06	
Fudenberg et al. (2012)	0.976	$\gg$	0.473	$\approx$	0.509	$\approx$	0.2		0.854	$\gg$	0.5	$\approx$	0.328	$\gg$	0.077	
Kagel and Schley (2013)	0.969	$\gg$	0.218	$\approx$	0.293	$>$	0.098		0.868	$\gg$	0.332	$\approx$	0.394	$\gg$	0.02	
Sherstyuk et al. (2013)	0.974	$\gg$	0.465	$\approx$	0.486	$\gg$	0.107		0.952	$\gg$	0.505	$\approx$	0.369	$\gg$	0.072	
Pooled	0.975	$\gg$	0.282	$\ll$	0.351	$\gg$	0.07		0.908	$\gg$	0.378	$\approx$	0.404	$\gg$	0.033	

Note: Relation signs, bootstrap procedure, and derived p-values are exactly as above, see Table 2, with the obvious adaptation that the Holm-Bonferroni correction now applies to all eight tests per data set.

Table 36: Table 35 by treatments – Strategies as a function of behavior in $t-2$ (Grim scheme)

(a) First halves per session

Treatment	Equality p-value	Cooperation after $(d,c),(c,d),(d,d)$ in $t-2$ $\hat{\sigma}_{cc}$		$\hat{\sigma}_{cd}$		$\hat{\sigma}_{dc}$		$\hat{\sigma}_{dd}$	Cooperation after ../tex-r1/ext-grim-tab2.tex in $t-2$ $\hat{\sigma}_{cc}$		$\hat{\sigma}_{cd}$		$\hat{\sigma}_{dc}$		$\hat{\sigma}_{dd}$
Aoyagi and Frechette (2009)															
AF09–34	0	0.939	$\gg$	0.39	$\approx$	0.439	$\approx$	0.556	0.782	$\gg$	0.485	$\approx$	0.39	$>$	0.32
Blonski et al. (2011)															
BOS11–9	0.23	-		0.182		0.182		0.031							
BOS11–14	0.16	-		0.188		0.062		0.029							
BOS11–15	0.04	-		0.167		0		0.005							
BOS11–16	0	0.934	$\gg$	0.136	$\approx$	0.136	$\approx$	0.056	0.667	$\approx$	0.333	$\approx$	0.333	$>$	0.076
BOS11–17	1	0.5	$\approx$	0.231	$\approx$	0.462	$\approx$	0.115	NA	$\approx$	0.25	$\approx$	0.5	$\approx$	0.167
BOS11–26	0.005	0.857	$>$	0.258	$\approx$	0.097	$\approx$	0.07	0.5	$\approx$	0.2	$\approx$	0.35	$\gg$	0.02
BOS11–27	0.18	0.875	$\approx$	0.556	$\approx$	0.333	$\approx$	0.091	1	$\approx$	0.1	$\approx$	0.1	$\approx$	0.044
BOS11–30	0.275	-		0		0		0.058							
BOS11–31	0	0.983	$\gg$	0.385	$\approx$	0.231	$\approx$	0.091	0.5	$\approx$	0.577	$\approx$	0.115	$\approx$	0.015
BOS11–All	0	0.903	$\gg$	0.248	$\approx$	0.174	$\gg$	0.045	0.714	$>$	0.318	$\approx$	0.216	$>$	0.031
Bruttel and Kamecke (2012)															
BK12–28	0	0.919	$\gg$	0.296	$\approx$	0.245	$\approx$	0.179	0.833	$\gg$	0.278	$\approx$	0.213	$\gg$	0.071
Dal Bó (2005)															
D05–18	0	0.821	$\gg$	0.208	$\approx$	0.25	$\approx$	0.091	0.75	$\approx$	0.273	$\approx$	0.364	$\approx$	0.118
D05–19	0	0.954	$\gg$	0.175	$\approx$	0.333	$\approx$	0.158	1	$\gg$	0.243	$\approx$	0.405	$\gg$	0.044
D05–All	0	0.926	$\gg$	0.184	$\approx$	0.31	$\approx$	0.143	0.889	$\gg$	0.254	$\approx$	0.39	$\gg$	0.074
Dal Bó and Fréchette (2011)															
DF11–6	0.059	0.667	$\approx$	0.294	$\approx$	0.235	$\gg$	0.038	0.917	$\gg$	0.375	$\approx$	0.35	$\gg$	0.034
DF11–7	0.002	0.632	$>$	0.254	$\approx$	0.254	$\gg$	0.089	0.786	$>$	0.391	$\approx$	0.266	$\gg$	0.029
DF11–8	0	0.979	$\gg$	0.446	$\approx$	0.28	$\gg$	0.105	0.923	$\gg$	0.361	$\approx$	0.222	$\gg$	0.06
DF11–22	0	0.922	$\gg$	0.34	$\approx$	0.381	$\gg$	0.06	0.833	$\gg$	0.279	$\approx$	0.338	$\gg$	0.048
DF11–23	0	0.976	$\gg$	0.448	$\approx$	0.321	$\gg$	0.16	0.859	$\gg$	0.325	$\approx$	0.462	$\gg$	0.054
DF11–24	0	0.967	$\gg$	0.228	$\approx$	0.366	$>$	0.135	0.813	$\gg$	0.308	$\approx$	0.436	$\gg$	0.107
DF11–All	0	0.961	$\gg$	0.342	$\approx$	0.307	$\gg$	0.081	0.849	$\gg$	0.324	$\approx$	0.364	$\gg$	0.054
Dal Bó and Fréchette (2015)															
DF15–4	0.017	0.571	$>$	0.073	$\approx$	0.268	$>$	0.018	0.5	$\approx$	0.429	$\approx$	0.5	$\gg$	0.044
DF15–5	0	0.92	$\gg$	0.223	$\approx$	0.219	$\gg$	0.076	0.95	$\gg$	0.369	$\approx$	0.323	$\gg$	0.086
DF15–20	0	0.933	$\gg$	0.222	$\approx$	0.335	$\gg$	0.073	0.825	$\gg$	0.225	$\approx$	0.337	$\gg$	0.046
DF15–21	0	0.959	$\gg$	0.325	$\approx$	0.329	$\gg$	0.129	0.873	$\gg$	0.455	$>$	0.411	$\gg$	0.077
DF15–33	0	0.953	$\gg$	0.313	$\approx$	0.322	$\gg$	0.111	0.802	$\gg$	0.288	$\approx$	0.356	$\gg$	0.047
DF15–35	0	0.98	$\gg$	0.276	$\approx$	0.448	$\approx$	0.214	0.882	$\gg$	0.356	$\approx$	0.422	$\gg$	0.042
DF15–All	0	0.95	$\gg$	0.265	$\approx$	0.301	$\gg$	0.081	0.843	$\gg$	0.328	$\approx$	0.369	$\gg$	0.052
Dreber et al. (2008)															
DRFN08–10	0	0.885	$\gg$	0.143	$\approx$	0.13	$>$	0.031	1	$>$	0.333	$\approx$	0.167	$>$	0.018
DRFN08–11	0	0.914	$\gg$	0.167	$\approx$	0.318	$>$	0.091	1	$>$	0.375	$\approx$	0.225	$\gg$	0.043
DRFN08–All	0	0.901	$\gg$	0.154	$\approx$	0.217	$\gg$	0.062	1	$\gg$	0.359	$\approx$	0.203	$\gg$	0.031
Duffy and Ochs (2009)															
DO09–32	0	0.932	$\gg$	0.218	$\approx$	0.301	$\approx$	0.208	0.748	$\gg$	0.361	$\approx$	0.35	$\gg$	0.102
Fréchette and Yuksel (2017)															
FY17–25	0	0.942	$\gg$	0.132	$\approx$	0.245	$\gg$	0	1	$\approx$	0.182	$\approx$	0.364	$\approx$	0.111
Fudenberg et al. (2012)															
FRD12–29	0	0.985	$\gg$	0.429	$\approx$	0.408	$\gg$	0	0.921	$\gg$	0.377	$\approx$	0.443	$\gg$	0.068
Kagel and Schley (2013)															
KS13–12	0	0.947	$\gg$	0.236	$\approx$	0.288	$\gg$	0.133	0.763	$\gg$	0.298	$\approx$	0.305	$\gg$	0.042
Sherstyuk et al. (2013)															
STS13–13	0	0.953	$\gg$	0.312	$\approx$	0.395	$\gg$	0.172	0.875	$\gg$	0.343	$\approx$	0.349	$\gg$	0.107
Pooled	0	0.949	$\gg$	0.278	$\approx$	0.3	$\gg$	0.091	0.825	$\gg$	0.333	$\approx$	0.346	$\gg$	0.059

(b) Second halves per session

Treatment	Equality p-value	Cooperation after $(d,c),(c,d),(d,d)$ in $t-2$ $\hat{\sigma}_{cc}$		$\hat{\sigma}_{cd}$		$\hat{\sigma}_{dc}$		$\hat{\sigma}_{dd}$	Cooperation after ../tex-r1/ext-grim-tab3.tex in $t-2$ $\hat{\sigma}_{cc}$		$\hat{\sigma}_{cd}$		$\hat{\sigma}_{dc}$		$\hat{\sigma}_{dd}$
Aoyagi and Frechette (2009)															
AF09–34	0	0.965	$\gg$	0.438	$\approx$	0.625	$\approx$	0.333	0.846	$\gg$	0.371	$\approx$	0.443	$\approx$	0.378
Blonski et al. (2011)															
BOS11–9	0.006	0.917	$>$	0	$\approx$	0.154	$\approx$	0.021	NA	$\approx$	0.333	$\approx$	0.333	$\approx$	0
BOS11–14	0.025	-		0.2		0.4		0.013							
BOS11–15	0	-		0		0		0.002							
BOS11–16	0	0.855	$\gg$	0.12	$\approx$	0.24	$>$	0	0.5	$\approx$	0.6	$\approx$	0.2	$\approx$	0.03
BOS11–17	0	0.912	$\gg$	0.161	$\approx$	0.194	$\approx$	0.048	1	$\approx$	0.208	$\approx$	0.333	$\gg$	0.024
BOS11–26	0	0.955	$\gg$	0.171	$\approx$	0.195	$\gg$	0.022	1	$>$	0.316	$\approx$	0.211	$\gg$	0.033
BOS11–27	0.01	0.867	$>$	0.31	$\approx$	0.414	$>$	0.109	0.9	$\gg$	0.518	$\approx$	0.268	$>$	0.014
BOS11–30	0.099	0.75	$\approx$	0.083	$\approx$	0.083	$\approx$	0	1	$\approx$	0.333	$\approx$	0.25	$\approx$	0.022
BOS11–31	0.004	1	$>$	0.143	$\approx$	0.286	$\approx$	0.062	1	$>$	0.613	$\approx$	0.097	$>$	0.018
BOS11–All	0	0.922	$\gg$	0.157	$\approx$	0.232	$\gg$	0.027	0.941	$\gg$	0.425	$\approx$	0.23	$\gg$	0.019
Bruttel and Kamecke (2012)															
BK12–28	0	0.946	$\gg$	0.156	$\approx$	0.233	$\approx$	0.173	0.958	$\gg$	0.327	$\approx$	0.4	$\gg$	0.019
Dal Bó (2005)															
D05–18	0	0.85	$\gg$	0.227	$<$	0.523	$\approx$	0.194	0.9	$\gg$	0.325	$\approx$	0.425	$\gg$	0.076
D05–19	0	0.949	$\gg$	0.13	$\approx$	0.283	$\approx$	0.083	1	$\gg$	0.314	$\approx$	0.314	$\gg$	0.04
D05–All	0	0.918	$\gg$	0.178	$<$	0.4	$>$	0.131	0.937	$\gg$	0.32	$\approx$	0.373	$\gg$	0.052
Dal Bó and Fréchette (2011)															
DF11–6	0.006	1	$>$	0.267	$\approx$	0.378	$\gg$	0.031	1	$\gg$	0.442	$\approx$	0.581	$\gg$	0.012
DF11–7	0	0.903	$\gg$	0.36	$\approx$	0.346	$\gg$	0.12	0.95	$\gg$	0.4	$\approx$	0.389	$\gg$	0.042
DF11–8	0	1	$\gg$	0.395	$>$	0.163	$\gg$	0.047	1	$\gg$	0.453	$\approx$	0.266	$\gg$	0.02
DF11–22	0	0.971	$\gg$	0.462	$\approx$	0.387	$\gg$	0.056	0.903	$\gg$	0.265	$\approx$	0.426	$\gg$	0.018
DF11–23	0	0.974	$\gg$	0.387	$\approx$	0.6	$\approx$	0.314	0.98	$\gg$	0.425	$\approx$	0.397	$\gg$	0.036
DF11–24	0	0.984	$\gg$	0.192	$\approx$	0.25	$\approx$	1	0.9	$>$	0.471	$\approx$	0.559	$\gg$	0.073
DF11–All	0	0.981	$\gg$	0.373	$\approx$	0.323	$\gg$	0.077	0.95	$\gg$	0.38	$\approx$	0.416	$\gg$	0.025
Dal Bó and Fréchette (2015)															
DF15–4	0.034	0.75	$>$	0.059	$\approx$	0.176	$\approx$	0.007	1	$\approx$	0.091	$\approx$	0.545	$>$	0.024
DF15–5	0	0.981	$\gg$	0.226	$\approx$	0.218	$\gg$	0.031	0.846	$\gg$	0.411	$\approx$	0.274	$\gg$	0.041
DF15–20	0	0.958	$\gg$	0.348	$\approx$	0.402	$\gg$	0.069	0.889	$\gg$	0.255	$\approx$	0.333	$\gg$	0.02
DF15–21	0	0.981	$\gg$	0.234	$\approx$	0.288	$>$	0.133	0.929	$\gg$	0.424	$\approx$	0.348	$\gg$	0.058
DF15–33	0	0.981	$\gg$	0.273	$\ll$	0.517	$\gg$	0.077	0.911	$\gg$	0.34	$<$	0.557	$\gg$	0.028
DF15–35	0	0.986	$\gg$	0.362	$\approx$	0.569	$\approx$	0.375	0.887	$\gg$	0.533	$\approx$	0.358	$\gg$	0.032
DF15–All	0	0.98	$\gg$	0.264	$<$	0.366	$\gg$	0.058	0.904	$\gg$	0.369	$\approx$	0.44	$\gg$	0.031
Dreber et al. (2008)															
DRFN08–10	0	0.667	$\gg$	0.02	$\approx$	0.18	$\approx$	0	1	$\approx$	0.75	$\approx$	0.875	$\gg$	0.002
DRFN08–11	0	0.943	$\gg$	0.036	$\ll$	0.436	$>$	0.031	0.929	$\gg$	0.321	$\approx$	0.536	$\gg$	0.028
DRFN08–All	0	0.913	$\gg$	0.029	$\ll$	0.314	$\gg$	0.007	0.955	$\gg$	0.417	$\approx$	0.611	$\gg$	0.009
Duffy and Ochs (2009)															
DO09–32	0	0.981	$\gg$	0.362	$\approx$	0.433	$\approx$	0.226	0.889	$\gg$	0.369	$\approx$	0.368	$\gg$	0.077
Fréchette and Yuksel (2017)															
FY17–25	0	0.976	$\gg$	0.173	$\approx$	0.308	$\approx$	0.222	0.75	$>$	0.294	$\approx$	0.49	$\gg$	0.06
Fudenberg et al. (2012)															
FRD12–29	0	0.976	$\gg$	0.473	$\approx$	0.509	$\approx$	0.2	0.854	$\gg$	0.5	$\approx$	0.328	$\gg$	0.077
Kagel and Schley (2013)															
KS13–12	0	0.969	$\gg$	0.218	$\approx$	0.293	$>$	0.098	0.868	$\gg$	0.332	$\approx$	0.394	$\gg$	0.02
Sherstyuk et al. (2013)															
STS13–13	0	0.974	$\gg$	0.465	$\approx$	0.486	$\gg$	0.107	0.952	$\gg$	0.505	$\approx$	0.369	$\gg$	0.072
Pooled	0	0.975	$\gg$	0.282	$\ll$	0.351	$\gg$	0.07	0.908	$\gg$	0.378	$\approx$	0.404	$\gg$	0.033

Table 37: 1- and 2-memory SG behavior strategies versus best mixtures (by treatment) of 1- and 2-memory pure strategies (No switching) (ICL-BIC of the models, less is better and relation signs point toward better models)

	SG+ SG M2"General"		SG M2"General"		Semi-Grim		Best Pure		Pure M1+G2,T2		Pure M1
Specification											
# Models evaluated	1		1		1		5		1		1
# Pars estimated (by treatment)	7		3		3		32		5		3
# Parameters accounted for	7		3		3		3-8		5		3
First halves per session											
Aoyagi and Frechette (2009)	855.34	$\approx$	847.81	$\approx$	835.89	$\approx$	891.63	$\approx$	891.63	$\approx$	897.8
Blonski et al. (2011)	2337.47	$\gg$	2188.04	$\gg$	1089.36	$\ll$	1241.55	$\ll$	1367.4	$\gg$	1239.58
Bruttel and Kamecke (2012)	1025.99	$\approx$	1021.23	$\gg$	817.24	$\approx$	861.15	$\approx$	861.15	$\approx$	862.58
Dal Bó (2005)	968.96	$\gg$	907.92	$\gg$	653.33	$\approx$	678.73	$\ll$	713.82	$>$	678.73
Dal Bó and Fréchette (2011)	14795.82	$\approx$	14789.67	$\gg$	7282.65	$<$	7668.25	$\approx$	7725.56	$\approx$	7670.28
Dal Bó and Fréchette (2015)	13772.1	$\gg$	13479.92	$\gg$	8887.67	$\approx$	9096.12	$\ll$	9276.23	$\gg$	9116.67
Dreber et al. (2008)	1176.51	$\approx$	1165.17	$\gg$	838.33	$\approx$	875.56	$<$	905.5	$>$	875.56
Duffy and Ochs (2009)	1670.22	$\approx$	1650.03	$\gg$	1437.86	$\approx$	1449.33	$\approx$	1459.86	$\approx$	1449.33
Fréchette and Yuksel (2017)	393.16	$\approx$	372.41	$\approx$	335.07	$\approx$	319.92	$\ll$	344.74	$\gg$	319.92
Fudenberg et al. (2012)	466.79	$\approx$	452.21	$\gg$	398.38	$\ll$	474.56	$\approx$	474.56	$\approx$	479.33
Kagel and Schley (2013)	3526.46	$\approx$	3570.33	$\gg$	2912.53	$>$	2739.66	$\approx$	2760.67	$\approx$	2739.66
Sherstyuk et al. (2013)	1685.47	$\approx$	1691.71	$\gg$	1413	$\approx$	1421.45	$\approx$	1421.45	$\approx$	1428.6
Pooled	**42929.62**	$\gg$	**42318.82**	$\gg$	**27010.74**	$\ll$	**27851.71**	$\ll$	**28384.93**	$\gg$	**27867.48**
Second halves per session											
Aoyagi and Frechette (2009)	515.26	$>$	500.7	$\approx$	494.93	$\approx$	548.36	$\approx$	548.36	$\approx$	553.46
Blonski et al. (2011)	3075.21	$\gg$	2951.31	$\gg$	1441.28	$\ll$	1757.39	$\ll$	1863.15	$\gg$	1757.39
Bruttel and Kamecke (2012)	833.83	$\approx$	838.57	$\gg$	595.23	$\approx$	583.12	$\approx$	583.12	$\approx$	594.04
Dal Bó (2005)	1041.04	$\gg$	975.62	$\gg$	748.55	$\approx$	747.84	$\ll$	785.02	$\gg$	747.84
Dal Bó and Fréchette (2011)	13878.91	$\approx$	13949.42	$\gg$	6160.5	$\approx$	6250.91	$\approx$	6306.7	$\approx$	6430.56
Dal Bó and Fréchette (2015)	14391.56	$\approx$	14280.59	$\gg$	9015.88	$<$	9477.45	$\approx$	9544.21	$\approx$	9552.41
Dreber et al. (2008)	1118.6	$\approx$	1106.62	$\gg$	665.13	$\approx$	664.79	$\ll$	690.58	$\approx$	664.79
Duffy and Ochs (2009)	2016.24	$\approx$	1993.56	$\gg$	1794.26	$\ll$	2016.45	$\approx$	2016.45	$\approx$	2042.07
Fréchette and Yuksel (2017)	561.78	$>$	528.38	$\approx$	481.62	$\approx$	474.69	$\ll$	502.3	$>$	474.69
Fudenberg et al. (2012)	532.03	$\approx$	530.32	$>$	485.43	$<$	551	$\approx$	551	$\approx$	571.98
Kagel and Schley (2013)	2648.79	$\approx$	2676.25	$\gg$	2261.67	$\gg$	1919.9	$\approx$	1919.9	$\approx$	1971.47
Sherstyuk et al. (2013)	1248.54	$\approx$	1293.11	$>$	1087.07	$\approx$	1029.75	$\approx$	1029.75	$<$	1127.23
Pooled	**42117.12**	$>$	**41806.84**	$\gg$	**25340.99**	$<$	**26159.81**	$\ll$	**26522.89**	$\approx$	**26597.37**

Note: Relation signs, bootstrap procedure, and derived p-values are exactly as above, see Table 2. Pure M1 refers to TFT, Grim, and AD. G2 denotes Grim2. For definitions of the strategies see Table 12.

Table 38: Table 37 by treatments – 1- and 2-memory SG behavior strategies versus best mixtures (by treatment) of 1- and 2-memory pure strategies (No switching)

(a) First halves per session

	SG+ SG M2"General"		SG M2"General"		Semi-Grim		Best Pure		Pure M1+G2,T2		Pure M1
Specification											
# Models evaluated	1		1		1		5		1		1
# Pars estimated (by treatment)	7		3		3		32		5		3
# Parameters accounted for	7		3		3		3-8		5		3
AF09–34	855.34	≈	847.81	≈	835.89	≈	891.63	≈	891.63	≈	897.8
BOS11–9	228.81	≫	213.06	≫	82.26	≈	99.25	≪	112.53	≫	99.25
BOS11–14	260.4	≈	249.04	≫	87.7	≪	114.56	≪	127.38	≫	114.56
BOS11–15	273.29	>	256.51	≫	31.07	≪	72.06	≪	85.01	≫	72.06
BOS11–16	225	>	210.23	>	170.96	≈	176.25	<	188.95	>	176.25
BOS11–17	165.97	≫	148.89	>	119.9	≈	123.55	≪	139.81	≫	123.55
BOS11–26	509.26	>	483.9	≫	254.14	≈	275.47	≪	293.71	>	275.47
BOS11–27	227.67	>	217.47	≫	108.51	≈	116.34	≈	116.34	≈	116.68
BOS11–30	173.13	≫	156.27	≫	63.02	≈	64.25	≪	77.67	≫	64.25
BOS11–31	203.85	≈	202.6	>	141.75	≈	167.46	≈	175.92	>	167.46
BK12–28	1025.99	≈	1021.23	≫	817.24	≈	861.15	≈	861.15	≈	862.58
D05–18	305.64	≫	281.91	≫	224.25	≈	240.35	≪	263.72	≫	240.35
D05–19	658.37	≫	622.47	≫	426.95	≈	436.26	≈	446.55	≈	436.26
DF11–6	3212.26	≈	3201.72	≫	894.35	≈	929.98	≈	939.5	≈	929.98
DF11–7	3939.04	≈	3932.55	≫	1369.26	<	1498.06	≈	1530.47	≈	1498.06
DF11–8	3028.46	≈	3058.62	≫	1648.75	≈	1568.94	≈	1568.94	≈	1571.62
DF11–22	1830.58	≈	1822.78	≫	1042.26	≪	1226.5	≈	1236.62	≈	1226.5
DF11–23	1453.39	≈	1461.14	≫	1117.42	≈	1176.02	≈	1176.02	≈	1181.42
DF11–24	1294.39	≈	1285.94	>	1194.45	≈	1246.18	≈	1247.07	≈	1246.54
DF15–4	1783.05	≈	1749.99	≫	477.13	≈	494.11	≪	523.98	≫	494.11
DF15–5	3031.87	>	2953.98	≫	2194.38	≫	1845.7	≈	1889.41	≈	1845.7
DF15–20	2763.46	≈	2713.19	≫	1481.46	<	1642.7	≈	1670.78	>	1642.7
DF15–21	2460.51	≈	2433.58	≫	2111.65	≈	2058.02	≈	2058.02	≈	2079.98
DF15–33	3248.44	>	3172.27	≫	2256.38	≪	2613.59	<	2676.03	>	2613.59
DF15–35	443.95	≈	427.76	≫	349.17	<	423.09	≈	428.84	≈	423.09
DRFN08–10	569.36	≈	559.43	≫	382.91	≈	415.65	<	432.45	≈	415.65
DRFN08–11	602.25	≈	602.24	>	453.32	≈	457.81	≈	469.55	≈	457.81
DO09–32	1670.22	≈	1650.03	≫	1437.86	≈	1449.33	≈	1459.86	≈	1449.33
FY17–25	393.16	≈	372.41	≈	335.07	≈	319.92	≪	344.74	≫	319.92
FRD12–29	466.79	≈	452.21	≫	398.38	≪	474.56	≈	474.56	≈	479.33
KS13–12	3526.46	≈	3570.33	≫	2912.53	>	2739.66	≈	2760.67	≈	2739.66
STS13–13	1685.47	≈	1691.71	≫	1413	≈	1421.45	≈	1421.45	≈	1428.6
Aoyagi and Frechette (2009)	855.34	≈	847.81	≈	835.89	≈	891.63	≈	891.63	≈	897.8
Blonski et al. (2011)	2337.47	≫	2188.04	≫	1089.36	≪	1241.55	≪	1367.4	≫	1239.58
Bruttel and Kamecke (2012)	1025.99	≈	1021.23	≫	817.24	≈	861.15	≈	861.15	≈	862.58
Dal Bó (2005)	968.96	≫	907.92	≫	653.33	≈	678.73	≪	713.82	>	678.73
Dal Bó and Fréchette (2011)	14795.82	≈	14789.67	≫	7282.65	<	7668.25	≈	7725.56	≈	7670.28
Dal Bó and Fréchette (2015)	13772.1	≫	13479.92	≫	8887.67	≈	9096.12	≪	9276.23	≫	9116.67
Dreber et al. (2008)	1176.51	≈	1165.17	≫	838.33	≈	875.56	<	905.5	>	875.56
Duffy and Ochs (2009)	1670.22	≈	1650.03	≫	1437.86	≈	1449.33	≈	1459.86	≈	1449.33
Fréchette and Yuksel (2017)	393.16	≈	372.41	≈	335.07	≈	319.92	≪	344.74	≫	319.92
Fudenberg et al. (2012)	466.79	≈	452.21	≫	398.38	≪	474.56	≈	474.56	≈	479.33
Kagel and Schley (2013)	3526.46	≈	3570.33	≫	2912.53	>	2739.66	≈	2760.67	≈	2739.66
Sherstyuk et al. (2013)	1685.47	≈	1691.71	≫	1413	≈	1421.45	≈	1421.45	≈	1428.6
Pooled	**42929.62**	≫	**42318.82**	≫	**27010.74**	≪	**27851.71**	≪	**28384.93**	≫	**27867.48**

(b) Second halves per session

	SG+ SG M2"General"		SG M2"General"		Semi-Grim		Best Pure		Pure M1+G2,T2		Pure M1
Specification											
# Models evaluated	1		1		1		5		1		1
# Pars estimated (by treatment)	7		3		3		32		5		3
# Parameters accounted for	7		3		3		3-8		5		3
AF09–34	515.26	>	500.7	≈	494.93	≈	548.36	≈	548.36	≈	553.46
BOS11–9	262.26	≈	252.38	≫	92.34	≈	100.69	≪	111.7	≫	100.69
BOS11–14	316.89	>	301.99	≫	35.25	≪	74.75	≪	86.28	≫	74.75
BOS11–15	357.64	>	340.78	≫	11.86	≪	93.99	≪	106.45	≫	93.99
BOS11–16	206.46	>	194.41	≈	162.19	≈	166.7	<	179.74	>	166.7
BOS11–17	327	≈	313.44	≫	212.8	≈	232.04	≈	237.43	≈	232.04
BOS11–26	635.75	≈	617.18	≫	333.39	<	391.24	≈	404.11	>	391.24
BOS11–27	388.11	≈	384.8	>	262.85	≈	292.85	≈	296.91	≈	292.85
BOS11–30	246.9	>	231.15	≫	131.11	<	161.56	≈	173.63	≫	161.56
BOS11–31	264.1	≈	265.11	>	169.45	<	213.53	≈	216.82	≈	213.53
BK12–28	833.83	≈	838.57	≫	595.23	≈	583.12	≈	583.12	≈	594.04
D05–18	450.68	≫	424.87	≫	336.26	≈	353.97	≈	364.98	≈	353.97
D05–19	585.4	≫	547.21	≫	410.17	≈	391.74	≪	416.49	≫	391.74
DF11–6	3524.22	≈	3504.27	≫	610.65	<	807.87	≈	830.4	≈	807.87
DF11–7	4029.73	≈	4054.52	≫	1566.15	≈	1638.56	≈	1655.2	≈	1638.56
DF11–8	2783	≈	2835.26	≫	1570.36	≫	1172.76	≈	1172.76	<	1228.26
DF11–22	1884.84	≈	1904.65	≫	1031.86	≈	1173.61	≈	1173.61	≈	1229.35
DF11–23	1003.64	≈	1024.22	>	885.87	≈	792.63	≈	792.63	≈	866.55
DF11–24	615.76	≈	599.58	>	479.46	≪	643.83	≈	655.18	≈	643.83
DF15–4	1896.41	≈	1863.03	≫	384.53	≈	417.63	≪	441.51	≫	417.63
DF15–5	3236.45	≈	3202.05	≫	2172.99	≫	1735.14	≈	1764.47	≈	1735.14
DF15–20	2796.05	≈	2784.9	≫	1393.96	<	1646.2	≈	1646.2	≈	1663.73
DF15–21	2061.25	≈	2051.69	≫	1826.53	≈	1832.26	≈	1840.23	≈	1832.26
DF15–33	3542.83	≈	3534.1	≫	2550.95	≪	3002.88	≈	3002.88	≈	3026.08
DF15–35	817.74	≈	815.66	≫	669.42	≪	819.77	≈	819.77	≈	860.08
DRFN08–10	664.97	≈	653.29	≫	288.29	≈	315.7	<	330.73	≈	315.7
DRFN08–11	448.73	≈	449.83	≈	374.74	≈	346.99	≈	356.35	≈	346.99
DO09–32	2016.24	≈	1993.56	≫	1794.26	≪	2016.45	≈	2016.45	≈	2042.07
FY17–25	561.78	>	528.38	≈	481.62	≈	474.69	≪	502.3	>	474.69
FRD12–29	532.03	≈	530.32	>	485.43	<	551	≈	551	≈	571.98
KS13–12	2648.79	≈	2676.25	≫	2261.67	≫	1919.9	≈	1919.9	≈	1971.47
STS13–13	1248.54	≈	1293.11	>	1087.07	≈	1029.75	≈	1029.75	<	1127.23
Aoyagi and Frechette (2009)	515.26	>	500.7	≈	494.93	≈	548.36	≈	548.36	≈	553.46
Blonski et al. (2011)	3075.21	≫	2951.31	≫	1441.28	≪	1757.39	≪	1863.15	≫	1757.39
Bruttel and Kamecke (2012)	833.83	≈	838.57	≫	595.23	≈	583.12	≈	583.12	≈	594.04
Dal Bó (2005)	1041.04	≫	975.62	≫	748.55	≈	747.84	≪	785.02	≫	747.84
Dal Bó and Fréchette (2011)	13878.91	≈	13949.42	≫	6160.5	≈	6250.91	≈	6306.7	≈	6430.56
Dal Bó and Fréchette (2015)	14391.56	≈	14280.59	≫	9015.88	<	9477.45	≈	9544.21	≈	9552.41
Dreber et al. (2008)	1118.6	≈	1106.62	≫	665.13	≈	664.79	≪	690.58	≈	664.79
Duffy and Ochs (2009)	2016.24	≈	1993.56	≫	1794.26	≪	2016.45	≈	2016.45	≈	2042.07
Fréchette and Yuksel (2017)	561.78	>	528.38	≈	481.62	≈	474.69	≪	502.3	>	474.69
Fudenberg et al. (2012)	532.03	≈	530.32	>	485.43	<	551	≈	551	≈	571.98
Kagel and Schley (2013)	2648.79	≈	2676.25	≫	2261.67	≫	1919.9	≈	1919.9	≈	1971.47
Sherstyuk et al. (2013)	1248.54	≈	1293.11	>	1087.07	≈	1029.75	≈	1029.75	<	1127.23
Pooled	**42117.12**	>	**41806.84**	≫	**25340.99**	<	**26159.81**	≪	**26522.89**	≈	**26597.37**

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 14.

Table 39: 1- and 2-memory SG behavior strategies versus best mixtures (by treatment) of 1- and 2-memory pure strategies (Random switching) (ICL-BIC of the models, less is better and relation signs point toward better models)

	SG+SG M2"General"		SG M2"General"		Semi-Grim		Best Pure		Pure 1+G2,T2		Pure 1
Specification											
# Models evaluated	1		1		1		5		1		1
# Pars estimated (by treatment)	7		3		3		32		5		3
# Parameters accounted for	7		3		3		3-8		5		3
First halves per session											
Aoyagi and Frechette (2009)	846.9	$\approx$	846.43	$\approx$	835.89	$\approx$	862.48	$\approx$	862.48	$\approx$	891.2
Blonski et al. (2011)	1938.36	$\gg$	1807.07	$\gg$	1089.36	$<$	1141.74	$<$	1168.97	$\approx$	1147.53
Bruttel and Kamecke (2012)	986.56	$\approx$	969.46	$\gg$	817.24	$\approx$	837.88	$\approx$	837.88	$\approx$	857.62
Dal Bó (2005)	810.96	$\approx$	798.82	$\gg$	653.33	$\ll$	690.17	$\approx$	693.14	$\approx$	692.19
Dal Bó and Fréchette (2011)	9041.66	$>$	8900.69	$\gg$	7282.65	$\ll$	7638.05	$\approx$	7645.17	$<$	7825.68
Dal Bó and Fréchette (2015)	11458.55	$\gg$	11208.45	$\gg$	8887.67	$\ll$	9301.85	$\approx$	9306.74	$<$	9460.02
Dreber et al. (2008)	1104.74	$\approx$	1080.21	$\gg$	838.33	$\approx$	871.6	$\approx$	871.39	$\approx$	879
Duffy and Ochs (2009)	1613.97	$\approx$	1588.23	$\gg$	1437.86	$\approx$	1488.29	$\approx$	1488.29	$\approx$	1507.66
Fréchette and Yuksel (2017)	400.09	$\gg$	363.06	$>$	335.07	$<$	349.53	$\approx$	354.74	$\approx$	352.99
Fudenberg et al. (2012)	442.4	$\approx$	440.73	$>$	398.38	$<$	445.18	$\approx$	445.18	$\approx$	466.32
Kagel and Schley (2013)	3481.07	$\approx$	3490.59	$\gg$	2912.53	$\approx$	2979.94	$\approx$	2979.94	$\approx$	3059.61
Sherstyuk et al. (2013)	1626.21	$\approx$	1601.7	$\gg$	1413	$<$	1483.72	$\approx$	1483.75	$\approx$	1503.01
Pooled	**34006.81**	$\gg$	**33277.82**	$\gg$	**27010.74**	$\ll$	**28272.53**	$\approx$	**28320.03**	$\ll$	**28752.26**
Second halves per session											
Aoyagi and Frechette (2009)	498.98	$\approx$	498.38	$\approx$	494.93	$\approx$	521.74	$\approx$	531.13	$\approx$	537.76
Blonski et al. (2011)	2648.18	$\gg$	2535.71	$\gg$	1441.28	$\ll$	1609.58	$\approx$	1637.46	$>$	1613.11
Bruttel and Kamecke (2012)	802.05	$\approx$	798.92	$\gg$	595.23	$\approx$	620.35	$\approx$	620.35	$\approx$	632.85
Dal Bó (2005)	915.79	$\approx$	904.46	$\gg$	748.55	$\ll$	793.71	$\approx$	796.01	$\approx$	807.66
Dal Bó and Fréchette (2011)	8212.23	$\approx$	8185.18	$\gg$	6160.5	$\ll$	6655.07	$\approx$	6655.57	$\ll$	6955.42
Dal Bó and Fréchette (2015)	12150.58	$>$	12016.88	$\gg$	9015.88	$\ll$	9880.99	$\approx$	9886.61	$\ll$	10178.06
Dreber et al. (2008)	1002.39	$\approx$	994.93	$\gg$	665.13	$\approx$	694.37	$\approx$	699.01	$\approx$	694.37
Duffy and Ochs (2009)	1970.43	$\approx$	1973.39	$\gg$	1794.26	$\ll$	2010.69	$\approx$	2010.69	$\approx$	2083.26
Fréchette and Yuksel (2017)	560.37	$>$	526.45	$\approx$	481.62	$\approx$	500.54	$\approx$	503.62	$\approx$	500.54
Fudenberg et al. (2012)	514.64	$\approx$	510.99	$\approx$	485.43	$\ll$	546.56	$\approx$	548.09	$<$	591.31
Kagel and Schley (2013)	2680.42	$\approx$	2675.61	$\gg$	2261.67	$\approx$	2312.35	$\approx$	2312.35	$<$	2399.2
Sherstyuk et al. (2013)	1256.5	$\approx$	1261.23	$\gg$	1087.07	$<$	1167.39	$\approx$	1167.39	$<$	1245.55
Pooled	**33467.87**	$\gg$	**33064.51**	$\gg$	**25340.99**	$\ll$	**27480.48**	$\approx$	**27550.66**	$\ll$	**28348.53**

Note: Relation signs, bootstrap procedure, and derived p-values are exactly as above, see Table 2. Pure M1 refers to TFT, Grim, and AD. G2 denotes Grim2. For definitions of the strategies see Table 12.

Table 40: Table 39 by treatments – 1- and 2-memory SG behavior strategies versus best mixtures (by treatment) of 1- and 2-memory pure strategies (Random switching)

(a) First halves per session

	SG+SG M2"General"		SG M2"General"		Semi-Grim		Best Pure		Pure 1+G2,T2		Pure 1
Specification											
# Models evaluated	1		1		1		5		1		1
# Pars estimated (by treatment)	7		3		3		32		5		3
# Parameters accounted for	7		3		3		3-8		5		3
AF09–34	846.9	≈	846.43	≈	835.89	≈	862.48	≈	862.48	≈	891.2
BOS11–9	168.61	≫	153.78	≫	82.26	≈	83.96	≈	86.52	≈	83.96
BOS11–14	210.98	≈	207.95	≫	87.7	≈	90.01	≈	92.44	≈	90.01
BOS11–15	221.49	≫	204.63	≫	31.07	≈	32.69	≈	33.3	≈	32.69
BOS11–16	212.6	>	199.27	>	170.96	≈	176.09	≈	177.8	≈	176.09
BOS11–17	142.08	≫	125.22	≈	119.9	≈	118.75	≈	121.83	≈	118.75
BOS11–26	458.54	≫	427.81	≫	254.14	≈	259.54	≈	259.54	≈	267.51
BOS11–27	167.95	≈	164.96	>	108.51	≈	109.62	≈	109.62	≈	113.56
BOS11–30	90.04	≫	73.18	>	63.02	≈	65.48	≈	68.48	≈	65.48
BOS11–31	195.97	≈	200.2	>	141.75	≈	169.35	≈	169.35	≈	169.44
BK12–28	986.56	≈	969.46	≫	817.24	≈	837.88	≈	837.88	≈	857.62
D05–18	262.83	≈	256.79	>	224.25	<	238.03	≈	240.12	≈	238.03
D05–19	543.17	≈	538.48	≫	426.95	<	449.48	≈	449.48	≈	452.03
DF11–6	1030.71	≈	998.16	≫	894.35	<	951.32	≈	951.32	≈	965.47
DF11–7	1482.77	≈	1460.6	>	1369.26	≈	1396.1	≈	1396.1	≈	1405.07
DF11–8	2145.02	≈	2117.53	≫	1648.75	<	1707.55	≈	1707.55	≈	1768.61
DF11–22	1643.66	≈	1622.27	≫	1042.26	≪	1139.78	≈	1140.58	≈	1162.24
DF11–23	1388.55	≈	1392.28	≫	1117.42	≈	1185.74	≈	1185.74	≈	1246
DF11–24	1313.25	≈	1282.93	>	1194.45	≈	1227.04	≈	1236.96	≈	1262.13
DF15–4	550.04	≫	511.86	≈	477.13	≈	475.23	≈	477.75	≈	475.23
DF15–5	2545.46	≫	2459.44	≫	2194.38	≈	2214.8	≈	2214.8	≈	2235.3
DF15–20	2434.72	≫	2354.53	≫	1481.46	<	1548.72	≈	1548.72	≈	1572.04
DF15–21	2359.09	≈	2351.69	≫	2111.65	≪	2225.46	≈	2225.46	<	2291.78
DF15–33	3090.18	≈	3078.85	≫	2256.38	≪	2420.93	≈	2421.39	<	2469.21
DF15–35	438.24	≈	422.92	≫	349.17	<	389.46	≈	389.46	≈	398.96
DRFN08–10	538.54	≈	517.1	≫	382.91	≈	395.82	≈	395.82	≈	398.11
DRFN08–11	561.29	≈	559.61	>	453.32	≈	471.86	≈	472.06	≈	478.79
DO09–32	1613.97	≈	1588.23	≫	1437.86	≈	1488.29	≈	1488.29	≈	1507.66
FY17–25	400.09	≫	363.06	>	335.07	<	349.53	≈	354.74	≈	352.99
FRD12–29	442.4	≈	440.73	>	398.38	<	445.18	≈	445.18	≈	466.32
KS13–12	3481.07	≈	3490.59	≫	2912.53	≈	2979.94	≈	2979.94	≈	3059.61
STS13–13	1626.21	≈	1601.7	≫	1413	<	1483.72	≈	1483.75	≈	1503.01
Aoyagi and Frechette (2009)	846.9	≈	846.43	≈	835.89	≈	862.48	≈	862.48	≈	891.2
Blonski et al. (2011)	1938.36	≫	1807.07	≫	1089.36	<	1141.74	<	1168.97	≈	1147.53
Bruttel and Kamecke (2012)	986.56	≈	969.46	≫	817.24	≈	837.88	≈	837.88	≈	857.62
Dal Bó (2005)	810.96	≈	798.82	≫	653.33	≪	690.17	≈	693.14	≈	692.19
Dal Bó and Fréchette (2011)	9041.66	>	8900.69	≫	7282.65	≪	7638.05	≈	7645.17	<	7825.68
Dal Bó and Fréchette (2015)	11458.55	≫	11208.45	≫	8887.67	≪	9301.85	≈	9306.74	<	9460.02
Dreber et al. (2008)	1104.74	≈	1080.21	≫	838.33	≈	871.6	≈	871.39	≈	879
Duffy and Ochs (2009)	1613.97	≈	1588.23	≫	1437.86	≈	1488.29	≈	1488.29	≈	1507.66
Fréchette and Yuksel (2017)	400.09	≫	363.06	>	335.07	<	349.53	≈	354.74	≈	352.99
Fudenberg et al. (2012)	442.4	≈	440.73	>	398.38	<	445.18	≈	445.18	≈	466.32
Kagel and Schley (2013)	3481.07	≈	3490.59	≫	2912.53	≈	2979.94	≈	2979.94	≈	3059.61
Sherstyuk et al. (2013)	1626.21	≈	1601.7	≫	1413	<	1483.72	≈	1483.75	≈	1503.01
Pooled	**34006.81**	≫	**33277.82**	≫	**27010.74**	≪	**28272.53**	≈	**28320.03**	≪	**28752.26**

(b) Second halves per session

	SG+SG M2"General"		SG M2"General"		Semi-Grim		Best Pure		Pure 1+G2,T2		Pure 1
Specification											
# Models evaluated	1		1		1		5		1		1
# Pars estimated (by treatment)	7		3		3		32		5		3
# Parameters accounted for	7		3		3		3-8		5		3
AF09–34	498.98	≈	498.38	≈	494.93	≈	521.74	≈	531.13	≈	537.76
BOS11–9	213.11	≫	196.77	≫	92.34	≈	96.65	≈	99.64	≈	96.65
BOS11–14	59.94	≫	43.82	≈	35.25	<	40.83	<	43.83	≈	40.83
BOS11–15	334.66	>	317.8	≫	11.86	≈	15.52	≪	18.52	≫	15.52
BOS11–16	196.1	≫	180.84	≈	162.19	≈	166.4	≈	168.97	≈	166.4
BOS11–17	322.09	≈	307.49	≫	212.8	≈	224.02	≈	224.02	≈	227.57
BOS11–26	577.71	≈	570.16	≫	333.39	<	375.4	≈	375.72	≈	375.4
BOS11–27	375.08	≈	375.07	>	262.85	≈	304.34	≈	304.34	≈	308.69
BOS11–30	231.84	≈	228.65	≫	131.11	≈	141.22	≈	144.11	≈	141.22
BOS11–31	267.55	≈	265.05	>	169.45	≈	208.24	≈	208.24	≈	210.78
BK12–28	802.05	≈	798.92	≫	595.23	≈	620.35	≈	620.35	≈	632.85
D05–18	390.84	≈	385.35	>	336.26	<	360.28	≈	361.7	≈	360.28
D05–19	519.98	≈	515.56	≫	410.17	<	430.77	≈	430.77	<	445.25
DF11–6	855.84	≈	852.06	≫	610.65	<	719.86	≈	719.86	≈	737.29
DF11–7	1807.16	≈	1803.16	≫	1566.15	≪	1683.6	≈	1683.6	≈	1710.18
DF11–8	2192.09	≈	2193.11	≫	1570.36	<	1641.29	≈	1641.29	<	1736.37
DF11–22	1811.46	≈	1810.44	≫	1031.86	<	1120.33	≈	1120.33	<	1218.47
DF11–23	1024.7	≈	1020.39	>	885.87	≈	946.29	≈	948.73	≈	1000.66
DF11–24	483.28	≈	479.09	≈	479.46	<	514.84	≈	514.84	≈	536.29
DF15–4	475.49	≫	438.92	>	384.53	≈	403.13	≈	405.76	≈	403.13
DF15–5	2818.99	≫	2728.87	≫	2172.99	<	2253.42	≈	2253.42	≈	2284.97
DF15–20	2476.53	≈	2470	≫	1393.96	<	1532.28	≈	1532.28	≈	1606.34
DF15–21	2052.8	≈	2046.05	≫	1826.53	≪	1939.3	≈	1940.01	<	1995.85
DF15–33	3515.78	≈	3532.97	≫	2550.95	≪	2910.55	≈	2911.64	<	3012.62
DF15–35	770.16	≈	770.91	>	669.42	≪	814.34	≈	814.34	≈	857.65
DRFN08–10	565.43	≈	561.61	≫	288.29	≈	300.31	≈	301.48	≈	300.31
DRFN08–11	432.06	≈	429.82	>	374.74	≈	391.96	≈	394.02	≈	391.96
DO09–32	1970.43	≈	1973.39	≫	1794.26	≪	2010.69	≈	2010.69	≈	2083.26
FY17–25	560.37	>	526.45	≈	481.62	≈	500.54	≈	503.62	≈	500.54
FRD12–29	514.64	≈	510.99	≈	485.43	≪	546.56	≈	548.09	<	591.31
KS13–12	2680.42	≈	2675.61	≫	2261.67	≈	2312.35	≈	2312.35	<	2399.2
STS13–13	1256.5	≈	1261.23	≫	1087.07	<	1167.39	≈	1167.39	<	1245.55
Aoyagi and Frechette (2009)	498.98	≈	498.38	≈	494.93	≈	521.74	≈	531.13	≈	537.76
Blonski et al. (2011)	2648.18	≫	2535.71	≫	1441.28	≪	1609.58	≈	1637.46	>	1613.11
Bruttel and Kamecke (2012)	802.05	≈	798.92	≫	595.23	≈	620.35	≈	620.35	≈	632.85
Dal Bó (2005)	915.79	≈	904.46	≫	748.55	≪	793.71	≈	796.01	≈	807.66
Dal Bó and Fréchette (2011)	8212.23	≈	8185.18	≫	6160.5	≪	6655.07	≈	6655.57	≪	6955.42
Dal Bó and Fréchette (2015)	12150.58	>	12016.88	≫	9015.88	≪	9880.99	≈	9886.61	≪	10178.06
Dreber et al. (2008)	1002.39	≈	994.93	≫	665.13	≈	694.37	≈	699.01	≈	694.37
Duffy and Ochs (2009)	1970.43	≈	1973.39	≫	1794.26	≪	2010.69	≈	2010.69	≈	2083.26
Fréchette and Yuksel (2017)	560.37	>	526.45	≈	481.62	≈	500.54	≈	503.62	≈	500.54
Fudenberg et al. (2012)	514.64	≈	510.99	≈	485.43	≪	546.56	≈	548.09	<	591.31
Kagel and Schley (2013)	2680.42	≈	2675.61	≫	2261.67	≈	2312.35	≈	2312.35	<	2399.2
Sherstyuk et al. (2013)	1256.5	≈	1261.23	≫	1087.07	<	1167.39	≈	1167.39	<	1245.55
Pooled	**33467.87**	≫	**33064.51**	≫	**25340.99**	≪	**27480.48**	≈	**27550.66**	≪	**28348.53**

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 14.

Table 41: 1-memory or 2-memory Semi-Grim strategies, complexity of memory, mixtures of 1-memory and 2-memory SG (no switching) (ICL-BIC of the models, less is better and relation signs point toward better models)

	SG M2"General"		SG M2"Semi-Grim"		SG M2"Grim"		Semi-Grim		SG M1 + M2"Grim"		SG M1 + M2"General"
Specification											
# Models evaluated	1		1		1		1		1		1
# Pars estimated (by treatment)	5		4		3		3		5		7
# Parameters accounted for	5		4		3		3		5		7
First halves per session											
Aoyagi and Frechette (2009)	865.91	$\approx$	864.19	$\approx$	843.09	$\approx$	835.89	$<$	906.78	$\gg$	848.86
Blonski et al. (2011)	1421.49	$>$	1397.85	$>$	1375.08	$\gg$	1089.36	$\ll$	1555.38	$<$	1591.37
Bruttel and Kamecke (2012)	969.35	$\approx$	967.88	$\approx$	966.75	$\gg$	817.24	$\ll$	968.22	$\approx$	968.68
Dal Bó (2005)	798.82	$\approx$	794.23	$\approx$	791.74	$\gg$	653.33	$\ll$	939.56	$>$	849.06
Dal Bó and Fréchette (2011)	8512.04	$\approx$	8495.44	$\approx$	8479.3	$\gg$	7282.65	$\ll$	8625.51	$\approx$	8659.2
Dal Bó and Fréchette (2015)	11283.53	$\approx$	11282.42	$\approx$	11280.81	$\gg$	8887.67	$\ll$	11725.16	$\approx$	11671.5
Dreber et al. (2008)	1177.21	$\approx$	1173.36	$\approx$	1170.29	$\gg$	838.33	$\ll$	1199.59	$\approx$	1202.45
Duffy and Ochs (2009)	1588.23	$\approx$	1587.59	$\approx$	1586.14	$\gg$	1437.86	$\ll$	1610.06	$\approx$	1629.09
Fréchette and Yuksel (2017)	362.68	$\approx$	360.88	$\approx$	359.65	$\approx$	335.07	$<$	365.5	$\approx$	369.6
Fudenberg et al. (2012)	440.73	$\approx$	442.18	$\approx$	440.31	$>$	398.38	$\ll$	452.14	$\approx$	455.2
Kagel and Schley (2013)	3490.59	$\approx$	3488.45	$\approx$	3478.42	$\gg$	2912.53	$\ll$	3438.64	$\approx$	3435.84
Sherstyuk et al. (2013)	1601.7	$\approx$	1602.31	$\approx$	1600.41	$\gg$	1413	$\ll$	1596.66	$\approx$	1598.4
Pooled	**32694.65**	$\approx$	**32602.68**	$\approx$	**32481.42**	$\gg$	**27010.74**	$\ll$	**33565.57**	$\approx$	**33534.58**
Second halves per session											
Aoyagi and Frechette (2009)	498.38	$\approx$	496.59	$\approx$	495.48	$\approx$	494.93	$\approx$	501.5	$\approx$	503.4
Blonski et al. (2011)	2277.26	$>$	2253.69	$>$	2230.87	$\gg$	1441.28	$\ll$	2411.7	$\approx$	2458.66
Bruttel and Kamecke (2012)	1013.73	$\approx$	1011.93	$\approx$	1010.14	$\gg$	595.23	$\ll$	1038.68	$\approx$	1040.33
Dal Bó (2005)	904.41	$\approx$	900.53	$\approx$	902.58	$\gg$	748.55	$\ll$	969.26	$\approx$	951.14
Dal Bó and Fréchette (2011)	8322.63	$\approx$	8300.39	$\approx$	8283.81	$\gg$	6160.5	$\ll$	8428.76	$\approx$	8461.64
Dal Bó and Fréchette (2015)	14925.81	$\approx$	14915.3	$\approx$	14901.44	$\gg$	9015.88	$\ll$	15226.83	$\approx$	15283.12
Dreber et al. (2008)	827.1	$\approx$	824.75	$\approx$	820.93	$\gg$	665.13	$\ll$	843.61	$\approx$	848.54
Duffy and Ochs (2009)	1973.39	$\approx$	1971.08	$\approx$	1968.89	$\gg$	1794.26	$\ll$	1988.35	$\approx$	1988.91
Fréchette and Yuksel (2017)	526.45	$\approx$	527.58	$\approx$	525.79	$\approx$	481.62	$<$	546.95	$<$	559.74
Fudenberg et al. (2012)	510.99	$\approx$	509.24	$\approx$	507.45	$\approx$	485.43	$\approx$	507.15	$\approx$	509.35
Kagel and Schley (2013)	2675.61	$\approx$	2674.59	$\approx$	2673.64	$\gg$	2261.67	$\ll$	2637.36	$\approx$	2623.98
Sherstyuk et al. (2013)	1261.23	$\approx$	1260.86	$\approx$	1259.78	$\gg$	1087.07	$<$	1219.58	$\approx$	1228.3
Pooled	**35899.35**	$\approx$	**35792.43**	$\approx$	**35690.23**	$\gg$	**25340.99**	$\ll$	**36502.1**	$\approx$	**36712.43**

Note: Relation signs, bootstrap procedure, and derived p-values are exactly as above.

Table 42: Table 41 by treatments – 1-memory or 2-memory Semi-Grim strategies, complexity of memory, mixtures of 1-memory and 2-memory SG (no switching)

(a) First halves per session

	SG M2"General"		SG M2"Semi-Grim"		SG M2"Grim"		Semi-Grim		SG M1 + M2"Grim"		SG M1 + M2"General"
Specification											
# Models evaluated	1		1		1		1		1		1
# Pars estimated (by treatment)	5		4		3		3		5		7
# Parameters accounted for	5		4		3		3		5		7
AF09–34	865.91	≈	864.19	≈	843.09	≈	835.89	<	906.78	≫	848.86
BOS11–9	95.12	≈	93.62	≈	92.12	>	82.26	≪	108.85	≈	111.63
BOS11–14	99.31	≈	97.81	≈	96.32	≈	87.7	≪	112.69	≈	114.94
BOS11–15	37.29	≈	35.79	≈	34.3	≈	31.07	≪	50.93	<	53.92
BOS11–16	199.22	≈	197.93	≈	197.5	>	170.96	<	211.68	≈	203.65
BOS11–17	125.22	≈	123.72	≈	122.22	≈	119.9	≪	139.08	≈	142.08
BOS11–26	335.64	≈	333.79	≈	331.95	>	254.14	≪	363.18	≈	366.9
BOS11–27	135.45	≈	133.95	≈	132.45	>	108.51	≪	149.17	≈	152.18
BOS11–30	73.18	≈	71.68	≈	70.18	≈	63.02	≪	87.04	≈	90.04
BOS11–31	270.99	≈	269.49	≈	267.99	≫	141.75	≪	282.68	≈	285.92
BK12–28	969.35	≈	967.88	≈	966.75	≫	817.24	≪	968.22	≈	968.68
D05–18	256.79	≈	254.92	≈	253.12	>	224.25	≪	301.43	≈	288.42
D05–19	538.48	≈	536.47	≈	536.5	≫	426.95	≪	634.58	>	555.67
DF11–6	998.16	≈	996.27	≈	994.37	≫	894.35	≪	1027.52	≈	1032.2
DF11–7	1460.6	≈	1458.64	≈	1456.69	>	1369.26	≪	1494.99	≈	1499
DF11–8	1990.14	≈	1988.22	≈	1986.31	≫	1648.75	≪	2021.5	≈	2025.23
DF11–22	1367.84	≈	1365.95	≈	1364.06	≫	1042.26	≪	1398.28	≈	1401.94
DF11–23	1385.44	≈	1383.73	≈	1381.92	≫	1117.42	≪	1379.96	≈	1386.38
DF11–24	1282.93	≈	1281.09	≈	1279.8	>	1194.45	<	1276.35	≈	1276.77
DF15–4	511.86	≈	509.91	≈	507.95	≈	477.13	≪	546.28	≈	549.97
DF15–5	2459.45	≈	2458.01	≈	2456.99	≫	2194.38	≪	2557.48	≈	2547.91
DF15–20	1939.87	≈	1937.5	≈	1935.13	≫	1481.46	≪	2018.71	≈	2021.82
DF15–21	2351.69	≈	2365.28	≈	2379.17	≫	2111.65	≪	2462.49	≫	2387.65
DF15–33	3568.57	≈	3566.01	≈	3563.45	≫	2256.38	≪	3682.71	≈	3689.45
DF15–35	422.92	≈	422.38	≈	420.63	≫	349.17	≪	428.33	≈	433.87
DRFN08–10	614.11	≈	612.45	≈	610.78	≫	382.91	≪	633.52	≈	635.7
DRFN08–11	559.59	≈	558.11	≈	557.41	>	453.32	<	562.57	≈	561.85
DO09–32	1588.23	≈	1587.59	≈	1586.14	≫	1437.86	≪	1610.06	≈	1629.09
FY17–25	362.68	≈	360.88	≈	359.65	≈	335.07	<	365.5	≈	369.6
FRD12–29	440.73	≈	442.18	≈	440.31	>	398.38	≪	452.14	≈	455.2
KS13–12	3490.59	≈	3488.45	≈	3478.42	≫	2912.53	≪	3438.64	≈	3435.84
STS13–13	1601.7	≈	1602.31	≈	1600.41	≫	1413	≪	1596.66	≈	1598.4
Aoyagi and Frechette (2009)	865.91	≈	864.19	≈	843.09	≈	835.89	<	906.78	≫	848.86
Blonski et al. (2011)	1421.49	>	1397.85	>	1375.08	≫	1089.36	≪	1555.38	<	1591.37
Bruttel and Kamecke (2012)	969.35	≈	967.88	≈	966.75	≫	817.24	≪	968.22	≈	968.68
Dal Bó (2005)	798.82	≈	794.23	≈	791.74	≫	653.33	≪	939.56	>	849.06
Dal Bó and Fréchette (2011)	8512.04	≈	8495.44	≈	8479.3	≫	7282.65	≪	8625.51	≈	8659.2
Dal Bó and Fréchette (2015)	11283.53	≈	11282.42	≈	11280.81	≫	8887.67	≪	11725.16	≈	11671.5
Dreber et al. (2008)	1177.21	≈	1173.36	≈	1170.29	≫	838.33	≪	1199.59	≈	1202.45
Duffy and Ochs (2009)	1588.23	≈	1587.59	≈	1586.14	≫	1437.86	≪	1610.06	≈	1629.09
Fréchette and Yuksel (2017)	362.68	≈	360.88	≈	359.65	≈	335.07	<	365.5	≈	369.6
Fudenberg et al. (2012)	440.73	≈	442.18	≈	440.31	>	398.38	≪	452.14	≈	455.2
Kagel and Schley (2013)	3490.59	≈	3488.45	≈	3478.42	≫	2912.53	≪	3438.64	≈	3435.84
Sherstyuk et al. (2013)	1601.7	≈	1602.31	≈	1600.41	≫	1413	≪	1596.66	≈	1598.4
Pooled	**32694.65**	≈	**32602.68**	≈	**32481.42**	≫	**27010.74**	≪	**33565.57**	≈	**33534.58**

(b) Second halves per session

	SG M2"General"		SG M2"Semi-Grim"		SG M2"Grim"		Semi-Grim		SG M1 + M2"Grim"		SG M1 + M2"General"
Specification											
# Models evaluated	1		1		1		1		1		1
# Pars estimated (by treatment)	5		4		3		3		5		7
# Parameters accounted for	5		4		3		3		5		7
AF09–34	498.38	≈	496.59	≈	495.48	≈	494.93	≈	501.5	≈	503.4
BOS11–9	132.55	≈	131.06	≈	129.56	>	92.34	≪	144.91	≈	147.93
BOS11–14	43.82	≈	42.32	≈	40.82	≈	35.25	≪	57.68	<	60.61
BOS11–15	14.95	>	13.45	>	11.95	≈	11.86	≪	28.81	<	31.81
BOS11–16	180.41	≈	179.18	≈	178.71	≈	162.19	<	193.95	≈	196.03
BOS11–17	366.29	≈	364.79	≈	363.3	≫	212.8	≪	380.12	≈	383.08
BOS11–26	521.64	≈	519.8	≈	517.95	≫	333.39	≪	549.32	≈	552.16
BOS11–27	390.6	≈	389.1	≈	387.61	>	262.85	<	403.67	≈	407.41
BOS11–30	178.29	≈	176.8	≈	175.3	>	131.11	<	192.09	≈	194.6
BOS11–31	398.63	≈	397.13	≈	395.63	≫	169.45	≪	411.07	≈	414.92
BK12–28	1013.73	≈	1011.93	≈	1010.14	≫	595.23	≪	1038.68	≈	1040.33
D05–18	385.3	≈	383.77	≈	382.42	>	336.26	≪	409.23	≈	410.57
D05–19	515.56	≈	513.92	≈	518.04	≫	410.17	≪	556.49	>	535.61
DF11–6	852.06	≈	850.17	≈	848.28	≫	610.65	≪	882.56	≈	886.33
DF11–7	1803.16	≈	1801.21	≈	1799.25	≫	1566.15	≪	1837.54	≈	1841.44
DF11–8	2239.88	≈	2237.97	≈	2236.05	≫	1570.36	≪	2271.07	≈	2274.87
DF11–22	1895.06	≈	1893.17	≈	1891.28	≫	1031.86	≪	1924.73	≈	1927.87
DF11–23	1026.45	≈	1018.73	≈	1017.07	>	885.87	<	998.41	≈	999.98
DF11–24	479.09	≈	477.61	≈	475.73	≈	479.46	≈	487.52	≈	493.46
DF15–4	438.92	≈	436.97	≈	435.01	>	384.53	≪	473.58	≈	477.38
DF15–5	2723.35	≈	2723.7	≈	2723.94	≫	2172.99	≪	2795.2	≈	2819.55
DF15–20	2389.9	≈	2387.53	≈	2385.16	≫	1393.96	≪	2468.89	≈	2473.66
DF15–21	2046.05	≈	2044.63	≈	2042.29	≫	1826.53	≪	2052.9	≈	2051.25
DF15–33	6527.52	≈	6524.96	≈	6522.39	≫	2550.95	≪	6637.01	≈	6648.75
DF15–35	770.91	≈	774.19	≈	775.14	≫	669.42	<	770.1	≈	771.71
DRFN08–10	393.77	≈	392.11	≈	390.44	≫	288.29	≪	411.68	≈	415.01
DRFN08–11	429.82	≈	429.84	≈	428.39	>	374.74	≈	428.43	≈	428.63
DO09–32	1973.39	≈	1971.08	≈	1968.89	≫	1794.26	≪	1988.35	≈	1988.91
FY17–25	526.45	≈	527.58	≈	525.79	≈	481.62	<	546.95	<	559.74
FRD12–29	510.99	≈	509.24	≈	507.45	≈	485.43	≈	507.15	≈	509.35
KS13–12	2675.61	≈	2674.59	≈	2673.64	≫	2261.67	≪	2637.36	≈	2623.98
STS13–13	1261.23	≈	1260.86	≈	1259.78	≫	1087.07	<	1219.58	≈	1228.3
Aoyagi and Frechette (2009)	498.38	≈	496.59	≈	495.48	≈	494.93	≈	501.5	≈	503.4
Blonski et al. (2011)	2277.26	>	2253.69	>	2230.87	≫	1441.28	≪	2411.7	≈	2458.66
Bruttel and Kamecke (2012)	1013.73	≈	1011.93	≈	1010.14	≫	595.23	≪	1038.68	≈	1040.33
Dal Bó (2005)	904.41	≈	900.53	≈	902.58	≫	748.55	≪	969.26	≈	951.14
Dal Bó and Fréchette (2011)	8322.63	≈	8300.39	≈	8283.81	≫	6160.5	≪	8428.76	≈	8461.64
Dal Bó and Fréchette (2015)	14925.81	≈	14915.3	≈	14901.44	≫	9015.88	≪	15226.83	≈	15283.12
Dreber et al. (2008)	827.1	≈	824.75	≈	820.93	≫	665.13	≪	843.61	≈	848.54
Duffy and Ochs (2009)	1973.39	≈	1971.08	≈	1968.89	≫	1794.26	≪	1988.35	≈	1988.91
Fréchette and Yuksel (2017)	526.45	≈	527.58	≈	525.79	≈	481.62	<	546.95	<	559.74
Fudenberg et al. (2012)	510.99	≈	509.24	≈	507.45	≈	485.43	≈	507.15	≈	509.35
Kagel and Schley (2013)	2675.61	≈	2674.59	≈	2673.64	≫	2261.67	≪	2637.36	≈	2623.98
Sherstyuk et al. (2013)	1261.23	≈	1260.86	≈	1259.78	≫	1087.07	<	1219.58	≈	1228.3
Pooled	**35899.35**	≈	**35792.43**	≈	**35690.23**	≫	**25340.99**	≪	**36502.1**	≈	**36712.43**

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 14.

Table 43: Mixtures of 1- and 2-memory pure and generalized strategies (no switching)
(ICL-BIC of the models, less is better and relation signs point toward better models)

	Gen M2		Gen M1		Best Pure M2		Pure M1		+ G2, TFT2, T2		+ 2TFT
Specification											
# Models evaluated	1		1		5		1		1		1
# Pars estimated (by treatment)	9		6		32		3		6		7
# Parameters accounted for	9		6		3–8		3		6		7
First halves per session											
Aoyagi and Frechette (2009)	764.25	$\approx$	757.68	$\ll$	884.86	$\approx$	892.99	$\approx$	890.72	$\approx$	892.49
Blonski et al. (2011)	1167.82	$\approx$	1208.25	$>$	1105.96	$\approx$	1105.89	$\ll$	1169.27	$<$	1195.88
Bruttel and Kamecke (2012)	827.89	$\approx$	853.09	$\approx$	839.97	$\approx$	851.01	$\approx$	841.77	$\approx$	843.55
Dal Bó (2005)	667.03	$\approx$	655.4	$\approx$	653.05	$\approx$	653.05	$\approx$	667.82	$\approx$	672.66
Dal Bó and Fréchette (2011)	7378.08	$\approx$	7433.78	$\approx$	7391.89	$\approx$	7453.78	$\approx$	7410.56	$\approx$	7426.49
Dal Bó and Fréchette (2015)	8826.62	$\approx$	8852.04	$\approx$	8893.78	$\approx$	8946.72	$\approx$	8929.45	$\approx$	8959.61
Dreber et al. (2008)	888.62	$\approx$	876.1	$\approx$	863.47	$\approx$	863.47	$\approx$	875.14	$\approx$	879.91
Duffy and Ochs (2009)	1414.26	$\approx$	1407.43	$\approx$	1426.34	$\approx$	1446.74	$\approx$	1429.36	$\approx$	1440.65
Fréchette and Yuksel (2017)	322.84	$\approx$	324.71	$\approx$	317.35	$\approx$	317.35	$<$	330.66	$\approx$	334.41
Fudenberg et al. (2012)	433.05	$\approx$	432.32	$\approx$	463.4	$\approx$	469.22	$\approx$	465.31	$\approx$	467.27
Kagel and Schley (2013)	2710.64	$\approx$	2739.15	$\approx$	2730.66	$\approx$	2737.32	$\approx$	2733.03	$\approx$	2737.72
Sherstyuk et al. (2013)	1386.14	$\approx$	1369.48	$\approx$	1398.69	$\approx$	1416.84	$\approx$	1400.69	$\approx$	1403.5
Pooled	**27115.51**	$\approx$	**27128.29**	$\approx$	**27115.38**	$\approx$	**27263.8**	$\approx$	**27362.62**	$\approx$	**27509.48**
Second halves per session											
Aoyagi and Frechette (2009)	417.68	$\approx$	416.51	$\ll$	540.47	$\approx$	543.34	$\approx$	546.38	$\approx$	544.96
Blonski et al. (2011)	1601.27	$\approx$	1588.79	$\approx$	1564.48	$\approx$	1567.21	$\approx$	1614.81	$\approx$	1640.42
Bruttel and Kamecke (2012)	575.98	$\approx$	592.59	$\approx$	567.99	$\approx$	587.38	$\approx$	569.78	$\approx$	571.6
Dal Bó (2005)	739.07	$<$	756.94	$\approx$	741.2	$\approx$	741.2	$\approx$	756.26	$\approx$	761.39
Dal Bó and Fréchette (2011)	5926.01	$<$	6059.85	$\approx$	5960.78	$\ll$	6189.93	$>$	5983.61	$\approx$	5994.24
Dal Bó and Fréchette (2015)	8955.93	$<$	9139.62	$\approx$	9143.98	$<$	9333.86	$>$	9170.84	$\approx$	9204.77
Dreber et al. (2008)	645.2	$\approx$	656.58	$\approx$	648.55	$\approx$	648.55	$<$	660.03	$\approx$	663.65
Duffy and Ochs (2009)	1888.67	$\approx$	1914.18	$\approx$	2003.41	$\approx$	2034.56	$\approx$	2005.7	$\approx$	2009.16
Fréchette and Yuksel (2017)	444.26	$\approx$	438.55	$<$	464.23	$\approx$	464.23	$\approx$	472.21	$\approx$	474.13
Fudenberg et al. (2012)	477.91	$\approx$	514.87	$\approx$	534.47	$\approx$	562.1	$\approx$	536.37	$\approx$	537.09
Kagel and Schley (2013)	1806.93	$\ll$	1923.93	$>$	1830.26	$<$	1924.38	$>$	1832.61	$\approx$	1835.1
Sherstyuk et al. (2013)	1029.88	$<$	1249.12	$\gg$	1023.43	$<$	1109.62	$>$	1025.44	$\approx$	1027.45
Pooled	**24837.07**	$\ll$	**25470.38**	$\approx$	**25177.57**	$\ll$	**25815.79**	$\gg$	**25392.89**	$\approx$	**25519.27**

Note: Relation signs, bootstrap procedure, and derived p-values are exactly as above, see Table 2. Pure M1 refers to TFT, Grim, and AD. G2 denotes Grim2. For definitions of pure strategies see Table 12. Gen M1 refers to generalized versions of TFT, Grim, and AD with memory-1. "+ G2, TFT2, T2" adds those strategies to the set of "Pure M1". "+2TFT" adds this strategy on top of the former.

Table 44: Table 43 by treatments – Mixtures of 1- and 2-memory pure and generalized strategies (no switching)

(a) First halves per session

	Gen M2		Gen M1		Best Pure M2		Pure M1		+ G2, TFT2, T2		+ 2TFT
Specification											
# Models evaluated	1		1		5		1		1		1
# Pars estimated (by treatment)	9		6		32		3		6		7
# Parameters accounted for	9		6		3–8		3		6		7
AF09–34	764.25	≈	757.68	≪	884.86	≈	892.99	≈	890.72	≈	892.49
BOS11–9	86.93	≈	89.7	≈	85.2	≈	85.2	≈	89.33	≈	90.83
BOS11–14	103.76	≈	102.27	≈	97.73	≈	97.73	≈	102.22	≈	103.88
BOS11–15	41.19	≈	38.79	>	34.3	≈	34.3	≪	38.79	≈	40.29
BOS11–16	173.13	≈	169.08	≈	174.24	≈	174.24	≈	178.7	≈	180.69
BOS11–17	119.24	≈	115.07	≈	110.57	≈	110.57	≈	115.16	≈	116.89
BOS11–26	259.69	≈	262.42	≈	256.88	≈	256.88	≈	259.48	≈	261.42
BOS11–27	102.01	≈	107.7	≈	100.97	≈	103.2	≈	102.47	≈	103.97
BOS11–30	65.81	≫	60.42	>	56.77	≈	56.77	<	61.29	≈	64.59
BOS11–31	125.92	≪	202.72	≈	156.95	≈	156.95	≈	161.73	≈	163.21
BK12–28	827.89	≈	853.09	≈	839.97	≈	851.01	≈	841.77	≈	843.55
D05–18	246.67	≈	241.44	≈	235.84	≈	235.84	≈	242.03	≈	243.88
D05–19	413.98	≈	409.7	≈	415.08	≈	415.08	≈	421.53	≈	423.82
DF11–6	883.72	≈	881.9	≈	877.78	≈	885.43	≈	879.67	≈	881.56
DF11–7	1436.53	≈	1432.97	≈	1424.78	≈	1424.78	≈	1430.65	≈	1432.61
DF11–8	1503.83	≈	1543.89	≈	1501.88	≈	1538.15	≈	1502.59	≈	1504.5
DF11–22	1178.2	≈	1185.37	≈	1188.65	≈	1189.26	≈	1190.54	≈	1191.58
DF11–23	1137.85	≈	1155.41	≈	1148.16	≈	1166.13	≈	1150.31	≈	1152.13
DF11–24	1189.48	≈	1201.92	≈	1224.49	≈	1233.88	≈	1224.49	≈	1226.42
DF15–4	468.06	≈	462.19	≈	456.32	≈	456.32	≈	462.19	≈	464.14
DF15–5	1756.46	≈	1762.23	≈	1817.09	≈	1818.32	≈	1819.54	≈	1821.88
DF15–20	1586.29	≈	1594.81	≈	1585.91	≈	1592.93	≈	1588.28	≈	1594.48
DF15–21	2002.93	≈	2003.37	≈	2022.58	≈	2069.89	≈	2025.12	≈	2029.73
DF15–33	2558.85	≈	2563.96	≈	2575.64	≈	2575.64	≈	2585.4	≈	2592.83
DF15–35	401.54	≈	430.5	≈	411.07	≈	416.14	≈	413.93	≈	415.7
DRFN08–10	424.56	≈	416.07	≈	410.24	≈	410.24	≈	415.24	≈	416.93
DRFN08–11	457.75	≈	455.83	≈	451.13	≈	451.13	≈	455.7	≈	458.08
DO09–32	1414.26	≈	1407.43	≈	1426.34	≈	1446.74	≈	1429.36	≈	1440.65
FY17–25	322.84	≈	324.71	≈	317.35	≈	317.35	<	330.66	≈	334.41
FRD12–29	433.05	≈	432.32	≈	463.4	≈	469.22	≈	465.31	≈	467.27
KS13–12	2710.64	≈	2739.15	≈	2730.66	≈	2737.32	≈	2733.03	≈	2737.72
STS13–13	1386.14	≈	1369.48	≈	1398.69	≈	1416.84	≈	1400.69	≈	1403.5
Aoyagi and Frechette (2009)	764.25	≈	757.68	≪	884.86	≈	892.99	≈	890.72	≈	892.49
Blonski et al. (2011)	1167.82	≈	1208.25	>	1105.96	≈	1105.89	≪	1169.27	<	1195.88
Bruttel and Kamecke (2012)	827.89	≈	853.09	≈	839.97	≈	851.01	≈	841.77	≈	843.55
Dal Bó (2005)	667.03	≈	655.4	≈	653.05	≈	653.05	≈	667.82	≈	672.66
Dal Bó and Fréchette (2011)	7378.08	≈	7433.78	≈	7391.89	≈	7453.78	≈	7410.56	≈	7426.49
Dal Bó and Fréchette (2015)	8826.62	≈	8852.04	≈	8893.78	≈	8946.72	≈	8929.45	≈	8959.61
Dreber et al. (2008)	888.62	≈	876.1	≈	863.47	≈	863.47	≈	875.14	≈	879.91
Duffy and Ochs (2009)	1414.26	≈	1407.43	≈	1426.34	≈	1446.74	≈	1429.36	≈	1440.65
Fréchette and Yuksel (2017)	322.84	≈	324.71	≈	317.35	≈	317.35	<	330.66	≈	334.41
Fudenberg et al. (2012)	433.05	≈	432.32	≈	463.4	≈	469.22	≈	465.31	≈	467.27
Kagel and Schley (2013)	2710.64	≈	2739.15	≈	2730.66	≈	2737.32	≈	2733.03	≈	2737.72
Sherstyuk et al. (2013)	1386.14	≈	1369.48	≈	1398.69	≈	1416.84	≈	1400.69	≈	1403.5
Pooled	**27115.51**	≈	**27128.29**	≈	**27115.38**	≈	**27263.8**	≈	**27362.62**	≈	**27509.48**

(b) Second halves per session

	Gen M2		Gen M1		Best Pure M2		Pure M1		+ G2, TFT2, T2		+ 2TFT
Specification											
# Models evaluated	1		1		5		1		1		1
# Pars estimated (by treatment)	9		6		32		3		6		7
# Parameters accounted for	9		6		3–8		3		6		7
AF09–34	417.68	≈	416.51	≪	540.47	≈	543.34	≈	546.38	≈	544.96
BOS11–9	96.6	≈	92.33	>	84.22	≈	84.22	<	88.72	≈	90.22
BOS11–14	49.8	>	45.31	≈	40.82	≈	40.82	≪	45.31	≈	46.81
BOS11–15	24.51	≈	20.01	≈	15.52	≈	15.52	≈	20.01	≈	21.51
BOS11–16	173.27	≈	162.81	≈	157.48	≈	157.48	≈	161.97	≈	163.73
BOS11–17	240.94	≈	234.24	≈	228.36	≈	229.75	≈	229.86	≈	231.36
BOS11–26	374.59	≈	375.35	≈	374.79	≈	375.99	≈	376.63	≈	379.73
BOS11–27	243.46	≈	290.73	≈	281.24	≈	286.24	≈	282.74	≈	284.24
BOS11–30	147.13	≈	148.34	≈	146.49	≈	146.49	≈	150.98	≈	152.47
BOS11–31	160.84	≈	159.57	≈	196.99	≈	200.65	≈	198.49	≈	200.23
BK12–28	575.98	≈	592.59	≈	567.99	≈	587.38	≈	569.78	≈	571.6
D05–18	346.26	≈	356.38	≈	350.59	≈	350.59	≈	354.41	≈	356.81
D05–19	386.44	≈	396.31	≈	388.48	≈	388.48	≈	397.6	≈	399.62
DF11–6	755.12	≈	752.32	≈	747.77	≈	747.77	≈	753.45	≈	755.34
DF11–7	1571.64	≈	1591.36	≈	1566.58	≈	1585.49	≈	1568.54	≈	1570.53
DF11–8	1140.35	<	1223.57	≈	1153.72	<	1217.82	≈	1155.64	≈	1157.54
DF11–22	1171.84	≈	1224.33	≈	1152.14	≈	1218.65	≈	1154.03	≈	1153.84
DF11–23	776.24	≈	785.37	≈	782.51	≈	863.26	>	786.34	≈	782.51
DF11–24	462.37	≈	450.61	<	530.97	≈	540.78	≈	533.31	≈	536.77
DF15–4	352.57	≈	347.64	≈	342.05	≈	342.05	≈	346.85	≈	348.81
DF15–5	1688.11	≈	1688.45	≈	1712.9	≈	1722.43	≈	1715.36	≈	1723.94
DF15–20	1563.94	≈	1628.58	≈	1582.66	≈	1622.15	≈	1585.03	≈	1587.38
DF15–21	1684.16	≈	1692.13	≈	1754.9	≈	1796.63	≈	1761.1	≈	1769.78
DF15–33	2856.1	<	2973.69	≈	2935.81	≈	2990.83	≈	2936.64	≈	2941.36
DF15–35	758.55	≈	774.14	≈	789.09	≈	842.27	>	790.87	≈	792.66
DRFN08–10	302.05	≈	303.34	≈	301.08	≈	301.08	≈	306.08	≈	307.74
DRFN08–11	336.84	≈	349.04	≈	345.37	≈	345.37	≈	349.75	≈	351
DO09–32	1888.67	≈	1914.18	≈	2003.41	≈	2034.56	≈	2005.7	≈	2009.16
FY17–25	444.26	≈	438.55	<	464.23	≈	464.23	≈	472.21	≈	474.13
FRD12–29	477.91	≈	514.87	≈	534.47	≈	562.1	≈	536.37	≈	537.09
KS13–12	1806.93	≪	1923.93	>	1830.26	<	1924.38	>	1832.61	≈	1835.1
STS13–13	1029.88	<	1249.12	≫	1023.43	<	1109.62	>	1025.44	≈	1027.45
Aoyagi and Frechette (2009)	417.68	≈	416.51	≪	540.47	≈	543.34	≈	546.38	≈	544.96
Blonski et al. (2011)	1601.27	≈	1588.79	≈	1564.48	≈	1567.21	≈	1614.81	≈	1640.42
Bruttel and Kamecke (2012)	575.98	≈	592.59	≈	567.99	≈	587.38	≈	569.78	≈	571.6
Dal Bó (2005)	739.07	<	756.94	≈	741.2	≈	741.2	≈	756.26	≈	761.39
Dal Bó and Fréchette (2011)	5926.01	<	6059.85	≈	5960.78	≪	6189.93	>	5983.61	≈	5994.24
Dal Bó and Fréchette (2015)	8955.93	<	9139.62	≈	9143.98	<	9333.86	>	9170.84	≈	9204.77
Dreber et al. (2008)	645.2	≈	656.58	≈	648.55	≈	648.55	<	660.03	≈	663.65
Duffy and Ochs (2009)	1888.67	≈	1914.18	≈	2003.41	≈	2034.56	≈	2005.7	≈	2009.16
Fréchette and Yuksel (2017)	444.26	≈	438.55	<	464.23	≈	464.23	≈	472.21	≈	474.13
Fudenberg et al. (2012)	477.91	≈	514.87	≈	534.47	≈	562.1	≈	536.37	≈	537.09
Kagel and Schley (2013)	1806.93	≪	1923.93	>	1830.26	<	1924.38	>	1832.61	≈	1835.1
Sherstyuk et al. (2013)	1029.88	<	1249.12	≫	1023.43	<	1109.62	>	1025.44	≈	1027.45
Pooled	**24837.07**	≪	**25470.38**	≈	**25177.57**	≪	**25815.79**	≫	**25392.89**	≈	**25519.27**

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 14.

Table 45: Comparison of 1- and 2-memory Semi-Grim with two and three parameters, pure and generalized strategies (no switching, Grim scheme) (ICL-BIC of the models, less is better and relation signs point toward better models)

	SGs M2"General"		SGs M2 "Grim"		Semi-Grim		Gen M2"Grim"		Gen M1		Best Pure M2
Specification											
# Models evaluated	1		1		1		1		1		5
# Pars estimated (by treatment)	5		3		3		9		6		32
# Parameters accounted for	5		3		3		9		6		3–8
First halves per session											
Aoyagi and Frechette (2009)	865.91	$>$	843.09	$\approx$	835.89	$>$	764.25	$\approx$	757.68	$\ll$	884.86
Blonski et al. (2011)	1421.49	$\gg$	1375.08	$\gg$	1089.36	$<$	1167.82	$\approx$	1208.25	$>$	1105.96
Bruttel and Kamecke (2012)	969.35	$\approx$	966.75	$\gg$	817.24	$\approx$	827.89	$\approx$	853.09	$\approx$	839.97
Dal Bó (2005)	798.82	$\approx$	791.74	$\gg$	653.33	$\approx$	667.03	$\approx$	655.4	$\approx$	653.05
Dal Bó and Fréchette (2011)	8512.04	$\approx$	8479.3	$\gg$	7282.65	$\approx$	7378.08	$\approx$	7433.78	$\approx$	7391.89
Dal Bó and Fréchette (2015)	11283.53	$\approx$	11280.81	$\gg$	8887.67	$\approx$	8826.62	$\approx$	8852.04	$\approx$	8893.78
Dreber et al. (2008)	1177.21	$\approx$	1170.29	$\gg$	838.33	$\approx$	888.62	$\approx$	876.1	$\approx$	863.47
Duffy and Ochs (2009)	1588.23	$\approx$	1586.14	$\gg$	1437.86	$\approx$	1414.26	$\approx$	1407.43	$\approx$	1426.34
Fréchette and Yuksel (2017)	362.68	$\approx$	359.65	$\approx$	335.07	$\approx$	322.84	$\approx$	324.71	$\approx$	317.35
Fudenberg et al. (2012)	440.73	$\approx$	440.31	$>$	398.38	$\approx$	433.05	$\approx$	432.32	$\approx$	463.4
Kagel and Schley (2013)	3490.59	$\approx$	3478.42	$\gg$	2912.53	$>$	2710.64	$\approx$	2739.15	$\approx$	2730.66
Sherstyuk et al. (2013)	1601.7	$\approx$	1600.41	$\gg$	1413	$\approx$	1386.14	$\approx$	1369.48	$\approx$	1398.69
Pooled	**32694.65**	$>$	**32481.42**	$\gg$	**27010.74**	$\approx$	**27115.51**	$\approx$	**27128.29**	$\approx$	**27115.38**
Second halves per session											
Aoyagi and Frechette (2009)	498.38	$\approx$	495.48	$\approx$	494.93	$>$	417.68	$\approx$	416.51	$\ll$	540.47
Blonski et al. (2011)	2277.26	$\gg$	2230.87	$\gg$	1441.28	$<$	1601.27	$\approx$	1588.79	$\approx$	1564.48
Bruttel and Kamecke (2012)	1013.73	$\approx$	1010.14	$\gg$	595.23	$\approx$	575.98	$\approx$	592.59	$\approx$	567.99
Dal Bó (2005)	904.41	$\approx$	902.58	$\gg$	748.55	$\approx$	739.07	$<$	756.94	$\approx$	741.2
Dal Bó and Fréchette (2011)	8322.63	$\approx$	8283.81	$\gg$	6160.5	$\approx$	5926.01	$<$	6059.85	$\approx$	5960.78
Dal Bó and Fréchette (2015)	14925.81	$\approx$	14901.44	$\gg$	9015.88	$\approx$	8955.93	$<$	9139.62	$\approx$	9143.98
Dreber et al. (2008)	827.1	$\approx$	820.93	$\gg$	665.13	$\approx$	645.2	$\approx$	656.58	$\approx$	648.55
Duffy and Ochs (2009)	1973.39	$\approx$	1968.89	$\gg$	1794.26	$\approx$	1888.67	$\approx$	1914.18	$\approx$	2003.41
Fréchette and Yuksel (2017)	526.45	$\approx$	525.79	$\approx$	481.62	$\approx$	444.26	$\approx$	438.55	$<$	464.23
Fudenberg et al. (2012)	510.99	$\approx$	507.45	$\approx$	485.43	$\approx$	477.91	$\approx$	514.87	$\approx$	534.47
Kagel and Schley (2013)	2675.61	$\approx$	2673.64	$\gg$	2261.67	$\gg$	1806.93	$\ll$	1923.93	$>$	1830.26
Sherstyuk et al. (2013)	1261.23	$\approx$	1259.78	$\gg$	1087.07	$\approx$	1029.88	$<$	1249.12	$\gg$	1023.43
Pooled	**35899.35**	$>$	**35690.23**	$\gg$	**25340.99**	$\approx$	**24837.07**	$\ll$	**25470.38**	$\approx$	**25177.57**

Note: Relation signs, bootstrap procedure, and derived p-values are exactly as above, see Table 2. Pure M1 refers to TFT, Grim, and AD. For definitions of pure strategies see Table 12. Gen M1 refers to generalized versions of TFT, Grim, and AD with memory-1. SGs refers to a two parameter version of SG $(1-\theta_1, \theta_2, \theta_2, \theta_1)$. "Gen M2" refers to memory-2 versions of the generalized strategies that allow parameters to depend on the prevalence of joint cooperation in $t-2$ (Grim Scheme).

Table 46: Comparison of 1- and 2-memory Semi-Grim, pure and generalized strategies (no switching, TFT scheme)
(ICL-BIC of the models, less is better and relation signs point toward better models)

	SGs M2"General"		SGs M2 "TFT"		Semi-Grim		Gen M2"TFT"		Gen M1		Best Pure M2
Specification											
# Models evaluated	1		1		1		1		1		5
# Pars estimated (by treatment)	5		3		3		9		6		32
# Parameters accounted for	5		3		3		9		6		3–8
First halves per session											
Aoyagi and Frechette (2009)	846.43	$\approx$	842.85	$\approx$	835.89	$>$	761.5	$\approx$	757.68	$\ll$	884.86
Blonski et al. (2011)	1806.09	$>$	1764.42	$\gg$	1089.36	$<$	1166.9	$\approx$	1208.25	$>$	1105.98
Bruttel and Kamecke (2012)	969.46	$\approx$	966.85	$\gg$	817.24	$\approx$	830.04	$\approx$	853.09	$\approx$	839.97
Dal Bó (2005)	798.82	$\approx$	792.19	$\gg$	653.33	$\approx$	670.93	$\approx$	655.4	$\approx$	653.05
Dal Bó and Fréchette (2011)	8766.46	$\approx$	8857.14	$\gg$	7282.65	$\approx$	7356.81	$\approx$	7433.78	$\approx$	7391.89
Dal Bó and Fréchette (2015)	11201.12	$\approx$	11195.82	$\gg$	8887.67	$\approx$	8772.73	$\approx$	8852.04	$\approx$	8893.78
Dreber et al. (2008)	1080.21	$\approx$	1074.01	$\gg$	838.33	$\approx$	885.14	$\approx$	876.1	$\approx$	863.47
Duffy and Ochs (2009)	1588.23	$\approx$	1589.78	$\gg$	1437.86	$\approx$	1408.4	$\approx$	1407.43	$\approx$	1426.34
Fréchette and Yuksel (2017)	362.68	$\approx$	359.82	$\approx$	335.07	$\approx$	317.71	$\approx$	324.71	$\approx$	317.35
Fudenberg et al. (2012)	440.73	$\approx$	438.77	$>$	398.38	$\approx$	434.18	$\approx$	432.32	$\approx$	463.4
Kagel and Schley (2013)	3482.28	$\approx$	3478.98	$\gg$	2912.53	$>$	2679.23	$\approx$	2739.15	$\approx$	2730.66
Sherstyuk et al. (2013)	1601.7	$\approx$	1599.88	$\gg$	1413	$\approx$	1361.19	$\approx$	1369.48	$\approx$	1398.69
Pooled	**33126.57**	$\approx$	**33069.94**	$\gg$	**27010.74**	$\approx$	**26973**	$\approx$	**27128.29**	$\approx$	**27115.39**
Second halves per session											
Aoyagi and Frechette (2009)	498.38	$\approx$	496.8	$\approx$	494.93	$>$	420.25	$\approx$	416.51	$\ll$	540.47
Blonski et al. (2011)	2534.05	$\approx$	2515.52	$\gg$	1441.28	$<$	1585.85	$\approx$	1588.79	$\approx$	1564.48
Bruttel and Kamecke (2012)	798.72	$\approx$	802.37	$\gg$	595.23	$\approx$	565.54	$\approx$	592.59	$\approx$	567.99
Dal Bó (2005)	904.46	$\approx$	903.44	$\gg$	748.55	$\approx$	736.51	$<$	756.94	$\approx$	741.2
Dal Bó and Fréchette (2011)	8180.26	$\approx$	8163	$\gg$	6160.5	$\approx$	5907.5	$<$	6059.85	$\approx$	5960.78
Dal Bó and Fréchette (2015)	12011.36	$\approx$	12036.08	$\gg$	9015.88	$\approx$	8931.28	$<$	9139.62	$\approx$	9143.98
Dreber et al. (2008)	994.91	$\approx$	990.52	$\gg$	665.13	$\approx$	640.2	$\approx$	656.58	$\approx$	648.55
Duffy and Ochs (2009)	1973.39	$\approx$	1969	$\gg$	1794.26	$\approx$	1866.23	$\approx$	1914.18	$\approx$	2003.41
Fréchette and Yuksel (2017)	526.45	$\approx$	524.04	$\approx$	481.62	$\approx$	442.91	$\approx$	438.55	$<$	464.23
Fudenberg et al. (2012)	510.99	$\approx$	508.14	$\approx$	485.43	$\approx$	503.36	$\approx$	514.87	$\approx$	534.47
Kagel and Schley (2013)	2675.61	$\approx$	2675.31	$\gg$	2261.67	$\gg$	1786.55	$\ll$	1923.93	$>$	1830.26
Sherstyuk et al. (2013)	1261.23	$\approx$	1260.14	$\gg$	1087.07	$\approx$	1015.79	$\ll$	1249.12	$\gg$	1023.43
Pooled	**33052.2**	$\approx$	**32953.79**	$\gg$	**25340.99**	$>$	**24730.23**	$\ll$	**25470.38**	$\approx$	**25177.57**

Note: Relation signs, bootstrap procedure, and derived p-values are exactly as above, see Table 2. Pure M1 refers to TFT, Grim, and AD. For definitions of pure strategies see Table 12. "Gen M1" refers to generalized versions of TFT, Grim, and AD with memory-1. SGs refers to a two parameter version of SG $(1-\theta_1, \theta_2, \theta_2, \theta_1)$. "Gen M2" refers to memory-2 versions of the generalized strategies that allow parameters to depend on opponent's behavior in $t-2$ (TFT Scheme).

Table 47: Examining all mixtures of Semi-Grim with pure or generalized pure strategies as secondary components

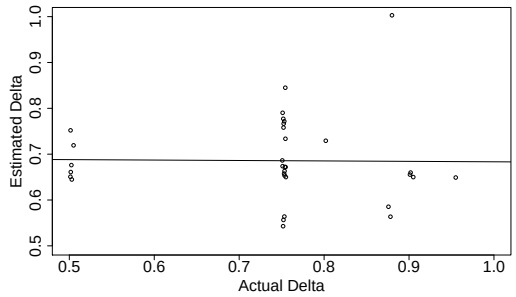
Component 1	First component is always Semi-Grim													
Component 2	Gen WSLS		Gen TFT		Gen Grim		Gen AD/AC		AD		Grim		TFT		WSLS
Specification															
# Models evaluated	1		1		1		1		1		1		1		1
# Pars estimated (by treatment)	5		5		5		5		4		4		4		4
# Parameters accounted for	5		5		5		5		4		4		4		4
First halves per session															
Aoyagi and Frechette (2009)	833.07	$\approx$	827.61	$>$	781.52	$\approx$	781.72	$\approx$	781.86	$\approx$	836.15	$\approx$	839.39	$\approx$	838.34
Blonski et al. (2011)	1104.67	$\ll$	1205.21	$\gg$	1078.98	$\approx$	1077.49	$\approx$	1069.28	$\approx$	1078.78	$\approx$	1084.91	$\approx$	1116.68
Bruttel and Kamecke (2012)	788.3	$\approx$	774.29	$\approx$	781.67	$\approx$	801.91	$\approx$	800.12	$\approx$	785.63	$\approx$	791.16	$\approx$	805.34
Dal Bó (2005)	626.71	$\approx$	615.03	$\approx$	618.81	$\approx$	633.79	$\approx$	629.17	$\approx$	622.09	$\approx$	620.42	$\ll$	665.94
Dal Bó and Fréchette (2011)	6792.73	$\approx$	6741.75	$\approx$	6717.88	$\approx$	6613.74	$\approx$	6597.93	$<$	6776.33	$\approx$	6768.78	$\ll$	7019.27
Dal Bó and Fréchette (2015)	8296.9	$>$	8219.32	$\approx$	8146.3	$\approx$	8032.68	$\approx$	8017.59	$\ll$	8282.13	$\approx$	8244.33	$\ll$	8578.97
Dreber et al. (2008)	783.18	$\approx$	778.8	$\approx$	780.25	$\approx$	786.29	$\approx$	782.37	$\approx$	779.27	$\approx$	801.73	$\approx$	840.26
Duffy and Ochs (2009)	1400.27	$\approx$	1392.67	$\approx$	1378.77	$\approx$	1375.28	$\approx$	1372.97	$\approx$	1398.16	$\approx$	1427.86	$\approx$	1368.08
Fréchette and Yuksel (2017)	296.99	$\approx$	288.62	$\approx$	291.8	$\approx$	301.54	$\approx$	299.62	$\approx$	295.03	$\approx$	301.61	$<$	341.67
Fudenberg et al. (2012)	407.24	$\approx$	393.68	$\approx$	396.2	$\approx$	382.94	$\approx$	381.01	$\approx$	408.96	$\approx$	391.66	$\approx$	403.85
Kagel and Schley (2013)	2707.66	$>$	2615.48	$\approx$	2659.78	$\approx$	2564.13	$\approx$	2561.76	$<$	2705.29	$\approx$	2737.96	$<$	2909.82
Sherstyuk et al. (2013)	1344.97	$\approx$	1288.49	$\approx$	1290.74	$\approx$	1305.81	$\approx$	1303.8	$\approx$	1342.96	$\approx$	1322.1	$<$	1417.88
Pooled	**25601.53**	$>$	**25359.79**	$>$	**25141.55**	$\approx$	**24876.16**	$\approx$	**24779.85**	$\ll$	**25493.16**	$\approx$	**25514.29**	$\ll$	**26488.47**
Second halves per session															
Aoyagi and Frechette (2009)	479.98	$>$	446.87	$>$	418.99	$\approx$	425.42	$\approx$	423.68	$\approx$	479.95	$\approx$	455.1	$\approx$	485.31
Blonski et al. (2011)	1439.96	$>$	1403.78	$\approx$	1398.73	$\approx$	1366.99	$\approx$	1346.79	$\approx$	1416.3	$\approx$	1397.1	$\approx$	1449.29
Bruttel and Kamecke (2012)	515.73	$\approx$	492.41	$\approx$	512.72	$\approx$	538.57	$\approx$	536.77	$\approx$	513.93	$\approx$	503.46	$\ll$	590.69
Dal Bó (2005)	693.22	$>$	673	$\approx$	697.25	$\approx$	710.27	$\approx$	699.05	$\approx$	688.6	$\approx$	707.82	$<$	761.12
Dal Bó and Fréchette (2011)	5253.2	$\approx$	5114.49	$\approx$	5119.08	$<$	5500.38	$>$	5128.69	$\approx$	5239.33	$\approx$	5098.31	$\ll$	5730.27
Dal Bó and Fréchette (2015)	7980.76	$\gg$	7744.75	$\approx$	7753.44	$\approx$	7873.59	$\approx$	7825.98	$\approx$	8016.72	$\approx$	7886.32	$\ll$	8545.26
Dreber et al. (2008)	568.76	$\approx$	551.89	$\approx$	565.89	$\approx$	593.75	$\approx$	589.84	$\approx$	564.85	$\approx$	573.93	$<$	672.96
Duffy and Ochs (2009)	1647.29	$\approx$	1715.88	$\approx$	1710.22	$\approx$	1661.28	$\approx$	1761.6	$\approx$	1747.49	$\approx$	1755.26	$\approx$	1795.33
Fréchette and Yuksel (2017)	464.38	$\approx$	437.79	$\approx$	431.82	$\approx$	425.29	$\approx$	423.34	$\approx$	462.62	$\approx$	457.07	$<$	487.74
Fudenberg et al. (2012)	463.31	$\approx$	473.75	$\approx$	481.89	$\approx$	470.44	$\approx$	452.6	$\approx$	461.37	$\approx$	483.91	$\approx$	487.78
Kagel and Schley (2013)	1902.37	$\gg$	1730.68	$\approx$	1791.78	$\approx$	1777.99	$\approx$	1775.62	$\approx$	1900	$\approx$	1881.46	$\ll$	2265.03
Sherstyuk et al. (2013)	1015.02	$\approx$	969.34	$\approx$	915.62	$\approx$	953.35	$\approx$	951.34	$\approx$	1013.01	$\approx$	986.2	$<$	1079.05
Pooled	**22642.83**	$\gg$	**21973.48**	$\approx$	**22016.28**	$<$	**22516.18**	$>$	**22097.67**	$<$	**22686.55**	$>$	**22368.31**	$\ll$	**24532.19**

Note: Relation signs, bootstrap procedure, and derived p-values are exactly as above, see Table 2. For definitions of pure strategies see Table 12. For definitions of generalized strategies see Section 3 main text.

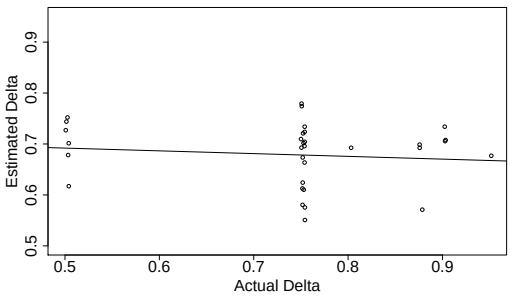
Appendix E: Robustness checks for Section 5

Figure 7: Relation of actual and estimated treatment parameters: Comparison of estimates based on regular and belief-free semi-grim MPEs (first halves of sessions)

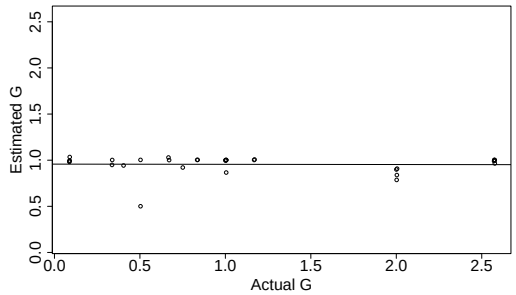
(a) RegSG: Actual vs. Estimated Delta



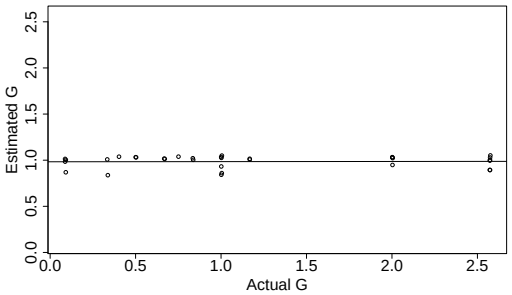
(b) BFSG: Actual vs. Estimated Delta



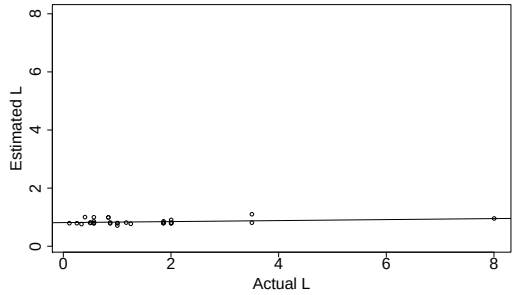
(c) RegSG: Actual vs. Estimated *g*



(d) BFSG: Actual vs. Estimated *g*



(e) RegSG: Actual vs. Estimated *l*



(f) BFSG: Actual vs. Estimated *l*

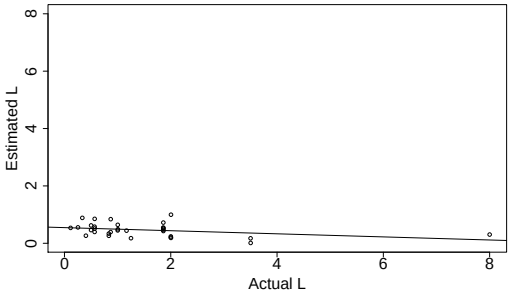
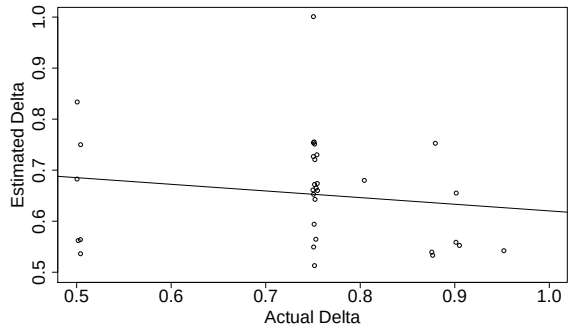
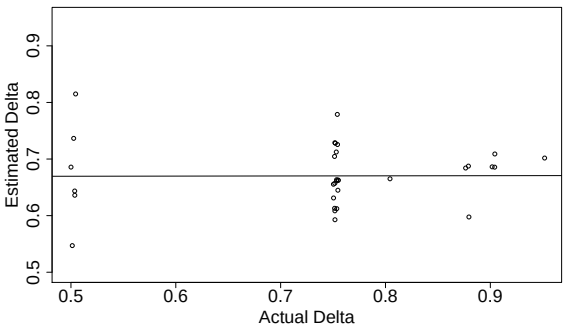


Figure 8: Relation of actual and estimated treatment parameters: Comparison of estimates based on regular and belief-free semi-grim MPEs (second halves of sessions)

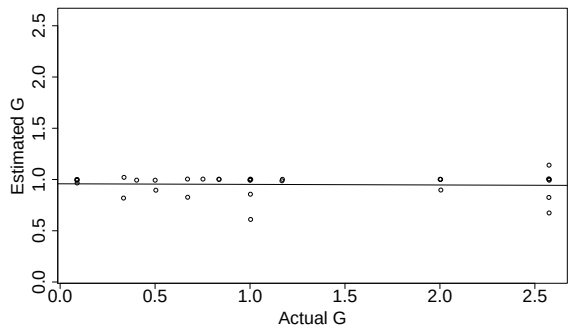
(a) RegSG: Actual vs. Estimated Delta



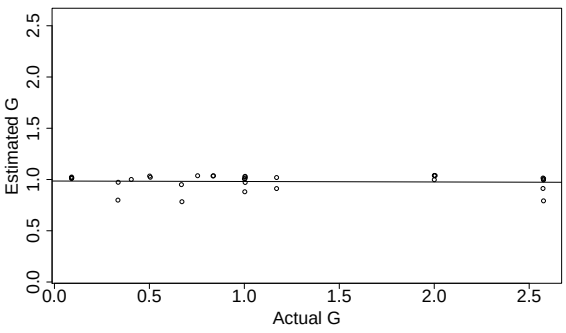
(b) BFSG: Actual vs. Estimated Delta



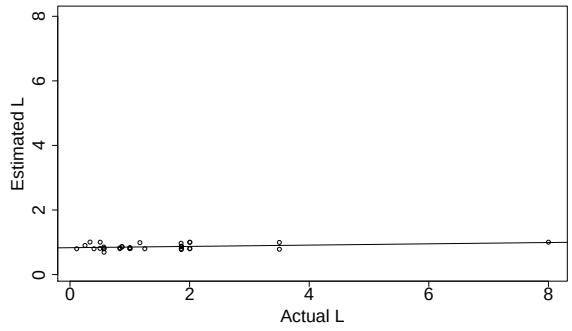
(c) RegSG: Actual vs. Estimated g



(d) BFSG: Actual vs. Estimated g



(e) RegSG: Actual vs. Estimated l



(f) BFSG: Actual vs. Estimated l

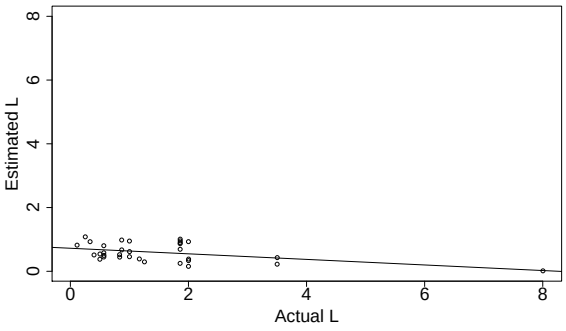


Table 48: Distance to Semi-Grim MPEs (first halves of sessions)

	Empirical		Closest Reg-SG MPE		Closest BF-SG MPE	
Treatment	SG-Strategy	Game (δ, g, l)	MAD	in game (δ, g, l)	MAD	in game (δ, g, l)
AF09–34	(0.91,0.41,0.41,0.09)	(0.9,0.33,0.11)	0.18	(0.65,1,0.79)	0	(0.7,0.84,0.54)
BOS11–9	(0.95,0.2,0.2,0.05)	(0.5,2,2)	0	(0.75,0.9,0.79)	0	(0.62,1.02,0.2)
BOS11–14	(0.99,0.12,0.12,0.01)	(0.75,0.5,3.5)	0	(0.84,1,0.81)	0	(0.55,1.03,0.17)
BOS11–15	(1,0.22,0.22,0)	(0.75,1,8)	0	(0.76,0.87,0.96)	0	(0.58,1.05,0.3)
BOS11–16	(0.95,0.18,0.18,0.05)	(0.75,0.75,1.25)	0.1	(0.77,0.92,0.77)	0	(0.61,1.04,0.17)
BOS11–17	(1,0.38,0.38,0)	(0.75,0.83,0.5)	0	(0.65,1,0.8)	0	(0.62,1,0.62)
BOS11–26	(0.98,0.17,0.17,0.02)	(0.75,2,2)	0.03	(0.79,0.78,0.81)	0.05	(0.58,1.03,0.24)
BOS11–27	(0.89,0.45,0.45,0.11)	(0.75,1,1)	0.23	(0.55,1,0.71)	0	(0.77,0.84,0.64)
BOS11–30	(1,0,0,0)	(0.88,0.5,3.5)	0	(1,0.5,1.1)	0	(0.57,1.03,0.01)
BOS11–31	(0.98,0.51,0.51,0.02)	(0.88,2,2)	0.05	(0.56,0.84,0.91)	0	(0.69,0.95,1)
BK12–28	(0.92,0.29,0.29,0.08)	(0.8,1.17,0.83)	0.17	(0.73,1,0.99)	0	(0.69,1.01,0.33)
D05–18	(0.86,0.29,0.29,0.14)	(0.75,1.17,0.83)	0.28	(0.73,1,0.99)	0	(0.77,1.01,0.26)
D05–19	(0.91,0.34,0.34,0.09)	(0.75,0.83,1.17)	0.17	(0.67,1,0.81)	0	(0.72,1.02,0.44)
DF11–6	(0.92,0.4,0.4,0.08)	(0.5,2.57,1.86)	0.16	(0.65,1,0.8)	0	(0.7,0.89,0.54)
DF11–7	(0.89,0.32,0.32,0.11)	(0.5,0.67,0.87)	0.21	(0.68,1,0.81)	0	(0.74,1.02,0.39)
DF11–8	(0.91,0.42,0.42,0.09)	(0.5,0.09,0.57)	0.19	(0.64,1,0.78)	0	(0.73,0.86,0.58)
DF11–22	(0.92,0.38,0.38,0.08)	(0.75,2.57,1.86)	0.17	(0.65,1,0.8)	0	(0.7,0.89,0.5)
DF11–23	(0.95,0.46,0.46,0.05)	(0.75,0.67,0.87)	0.1	(0.56,1.03,0.78)	0	(0.72,1.01,0.84)
DF11–24	(0.95,0.36,0.36,0.05)	(0.75,0.09,0.57)	0.1	(0.66,0.99,0.8)	0	(0.67,1,0.52)
DF15–4	(0.9,0.36,0.36,0.1)	(0.5,2.57,1.86)	0.21	(0.66,0.99,0.8)	0	(0.75,0.99,0.46)
DF15–5	(0.93,0.31,0.31,0.07)	(0.5,0.09,0.57)	0.14	(0.72,0.98,1)	0	(0.67,1.01,0.39)
DF15–20	(0.92,0.32,0.32,0.08)	(0.75,2.57,1.86)	0.16	(0.68,0.96,0.85)	0	(0.71,1.05,0.43)
DF15–21	(0.94,0.47,0.47,0.06)	(0.75,0.09,0.57)	0.12	(0.54,1.03,0.85)	0	(0.73,0.98,0.85)
DF15–33	(0.93,0.37,0.37,0.07)	(0.9,2.57,1.86)	0.14	(0.66,1,0.81)	0	(0.7,1,0.53)
DF15–35	(0.97,0.42,0.42,0.03)	(0.95,2.57,1.86)	0.06	(0.64,1,0.79)	0	(0.67,1.03,0.72)
DRFN08–10	(0.95,0.18,0.18,0.05)	(0.75,2,2)	0.11	(0.77,0.91,0.79)	0.02	(0.61,1.03,0.18)
DRFN08–11	(0.93,0.33,0.33,0.07)	(0.75,1,1)	0.14	(0.67,1,0.8)	0	(0.69,1.03,0.44)
DO09–32	(0.9,0.37,0.37,0.1)	(0.9,1,1)	0.2	(0.65,0.99,0.79)	0	(0.73,0.93,0.48)
FY17–25	(0.93,0.25,0.25,0.07)	(0.75,0.4,0.4)	0.15	(0.76,0.94,1)	0	(0.66,1.03,0.26)
FRD12–29	(0.97,0.47,0.47,0.03)	(0.88,0.33,0.33)	0.06	(0.58,0.95,0.76)	0	(0.7,1.01,0.88)
KS13–12	(0.93,0.33,0.33,0.07)	(0.75,1,0.5)	0.14	(0.67,1,0.81)	0	(0.69,1.02,0.46)
STS13–13	(0.92,0.41,0.41,0.08)	(0.75,1,0.25)	0.16	(0.65,1,0.79)	0	(0.7,0.86,0.55)
Means			0.123	(0.69,0.95,0.84)	0.002	(0.68,0.98,0.47)

Note: "SG-Strategy" is the SG-continuation strategy estimated in the "1.5× SG + AD" model

Table 49: Distance to Semi-Grim MPEs: closest equilibria (first halves of sessions)

Treatment	Empirical		Closest Reg-SG MPE	Closest BF-SG MPE
	SG-Strategy	Game	Strategy	Strategy
AF09–34	(0.91,0.41,0.41,0.09)	(0.9,0.33,0.11)	(1,0.41,0.41,0)	(0.91,0.41,0.41,0.09)
BOS11–9	(0.95,0.2,0.2,0.05)	(0.5,2,2)	(1,0.2,0.2,0)	(0.95,0.2,0.2,0.05)
BOS11–14	(0.99,0.12,0.12,0.01)	(0.75,0.5,3.5)	(1,0.12,0.12,0)	(1,0.15,0.15,0)
BOS11–15	(1,0.22,0.22,0)	(0.75,1,8)	(1,0.22,0.22,0)	(1,0.22,0.22,0)
BOS11–16	(0.95,0.18,0.18,0.05)	(0.75,0.75,1.25)	(1,0.18,0.18,0)	(0.95,0.18,0.18,0.05)
BOS11–17	(1,0.38,0.38,0)	(0.75,0.83,0.5)	(1,0.38,0.38,0)	(1,0.38,0.38,0)
BOS11–26	(0.98,0.17,0.17,0.02)	(0.75,2,2)	(1,0.17,0.17,0)	(0.98,0.2,0.2,0.02)
BOS11–27	(0.89,0.45,0.45,0.11)	(0.75,1,1)	(1,0.45,0.45,0)	(0.89,0.45,0.45,0.11)
BOS11–30	(1,0,0,0)	(0.88,0.5,3.5)	(1,0,0,0)	(0.95,0.06,0.06,0.05)
BOS11–31	(0.98,0.51,0.51,0.02)	(0.88,2,2)	(1,0.51,0.51,0)	(0.98,0.51,0.51,0.02)
BK12–28	(0.92,0.29,0.29,0.08)	(0.8,1.17,0.83)	(1,0.29,0.29,0)	(0.92,0.29,0.29,0.08)
D05–18	(0.86,0.29,0.29,0.14)	(0.75,1.17,0.83)	(1,0.29,0.29,0)	(0.86,0.29,0.29,0.14)
D05–19	(0.91,0.34,0.34,0.09)	(0.75,0.83,1.17)	(1,0.34,0.34,0)	(0.91,0.34,0.34,0.09)
DF11–6	(0.92,0.4,0.4,0.08)	(0.5,2.57,1.86)	(1,0.4,0.4,0)	(0.92,0.4,0.4,0.08)
DF11–7	(0.89,0.32,0.32,0.11)	(0.5,0.67,0.87)	(1,0.32,0.32,0)	(0.89,0.32,0.32,0.11)
DF11–8	(0.91,0.42,0.42,0.09)	(0.5,0.09,0.57)	(1,0.42,0.42,0)	(0.91,0.42,0.42,0.09)
DF11–22	(0.92,0.38,0.38,0.08)	(0.75,2.57,1.86)	(1,0.38,0.38,0)	(0.92,0.38,0.38,0.08)
DF11–23	(0.95,0.46,0.46,0.05)	(0.75,0.67,0.87)	(1,0.46,0.46,0)	(0.95,0.46,0.46,0.05)
DF11–24	(0.95,0.36,0.36,0.05)	(0.75,0.09,0.57)	(1,0.36,0.36,0)	(0.95,0.36,0.36,0.05)
DF15–4	(0.9,0.36,0.36,0.1)	(0.5,2.57,1.86)	(1,0.36,0.36,0)	(0.9,0.36,0.36,0.1)
DF15–5	(0.93,0.31,0.31,0.07)	(0.5,0.09,0.57)	(1,0.31,0.31,0)	(0.93,0.31,0.31,0.07)
DF15–20	(0.92,0.32,0.32,0.08)	(0.75,2.57,1.86)	(1,0.32,0.32,0)	(0.92,0.32,0.32,0.08)
DF15–21	(0.94,0.47,0.47,0.06)	(0.75,0.09,0.57)	(1,0.47,0.47,0)	(0.94,0.47,0.47,0.06)
DF15–33	(0.93,0.37,0.37,0.07)	(0.9,2.57,1.86)	(1,0.37,0.37,0)	(0.93,0.37,0.37,0.07)
DF15–35	(0.97,0.42,0.42,0.03)	(0.95,2.57,1.86)	(1,0.42,0.42,0)	(0.97,0.42,0.42,0.03)
DRFN08–10	(0.95,0.18,0.18,0.05)	(0.75,2,2)	(1,0.18,0.18,0)	(0.95,0.19,0.19,0.05)
DRFN08–11	(0.93,0.33,0.33,0.07)	(0.75,1,1)	(1,0.33,0.33,0)	(0.93,0.33,0.33,0.07)
DO09–32	(0.9,0.37,0.37,0.1)	(0.9,1,1)	(1,0.37,0.37,0)	(0.9,0.37,0.37,0.1)
FY17–25	(0.93,0.25,0.25,0.07)	(0.75,0.4,0.4)	(1,0.25,0.25,0)	(0.93,0.25,0.25,0.07)
FRD12–29	(0.97,0.47,0.47,0.03)	(0.88,0.33,0.33)	(1,0.47,0.47,0)	(0.97,0.47,0.47,0.03)
KS13–12	(0.93,0.33,0.33,0.07)	(0.75,1,0.5)	(1,0.33,0.33,0)	(0.93,0.33,0.33,0.07)
STS13–13	(0.92,0.41,0.41,0.08)	(0.75,1,0.25)	(1,0.41,0.41,0)	(0.92,0.41,0.41,0.08)

Note: "SG-Strategy" is the SG-continuation strategy estimated in the "1.5× SG + AD" model

Table 50: Distance to Semi-Grim MPEs (second halves of sessions)

	Empirical		Closest Reg-SG MPE		Closest BF-SG MPE	
Treatment	SG-Strategy	Game (δ, g, l)	MAD	in game (δ, g, l)	MAD	in game (δ, g, l)
AF09–34	(0.97,0.46,0.46,0.03)	(0.9,0.33,0.11)	0.06	(0.55,1.02,0.8)	0	(0.68,0.97,0.82)
BOS11–9	(1,0.13,0.13,0)	(0.5,2,2)	0	(0.83,1,0.8)	0	(0.54,1.03,0.15)
BOS11–14	(0.99,0.3,0.3,0.01)	(0.75,0.5,3.5)	0	(0.72,0.99,0.99)	0	(0.61,1.02,0.43)
BOS11–15	(1,0,0,0)	(0.75,1,8)	0	(1,0.99,1)	0	(0.59,1.03,0.01)
BOS11–16	(0.97,0.21,0.21,0.03)	(0.75,0.75,1.25)	0.07	(0.75,1,0.8)	0.06	(0.61,1.03,0.29)
BOS11–17	(0.95,0.26,0.26,0.05)	(0.75,0.83,0.5)	0.1	(0.75,1,1)	0.07	(0.64,1.04,0.38)
BOS11–26	(0.94,0.29,0.29,0.06)	(0.75,2,2)	0.13	(0.73,1,1)	0.03	(0.66,1.04,0.38)
BOS11–27	(0.95,0.5,0.5,0.05)	(0.75,1,1)	0.1	(0.51,0.86,0.83)	0	(0.73,0.97,0.95)
BOS11–30	(0.96,0.2,0.2,0.04)	(0.88,0.5,3.5)	0.08	(0.75,0.89,0.78)	0.02	(0.6,1.03,0.22)
BOS11–31	(0.98,0.48,0.48,0.02)	(0.88,2,2)	0.04	(0.54,0.89,0.81)	0	(0.69,0.99,0.93)
BK12–28	(0.95,0.32,0.32,0.05)	(0.8,1.17,0.83)	0.1	(0.68,0.98,0.82)	0	(0.66,1.02,0.44)
D05–18	(0.88,0.4,0.4,0.12)	(0.75,1.17,0.83)	0.24	(0.65,1,0.8)	0	(0.78,0.91,0.52)
D05–19	(0.95,0.3,0.3,0.05)	(0.75,0.83,1.17)	0.11	(0.72,1,0.98)	0	(0.66,1.03,0.38)
DF11–6	(0.94,0.55,0.55,0.06)	(0.5,2.57,1.86)	0.13	(0.56,0.67,0.97)	0	(0.73,0.79,1.01)
DF11–7	(0.86,0.47,0.47,0.14)	(0.5,0.67,0.87)	0.27	(0.53,1,0.86)	0	(0.81,0.78,0.67)
DF11–8	(0.97,0.45,0.45,0.03)	(0.5,0.09,0.57)	0.06	(0.56,0.99,0.69)	0	(0.68,1.01,0.8)
DF11–22	(0.96,0.47,0.47,0.04)	(0.75,2.57,1.86)	0.08	(0.55,1,0.81)	0	(0.7,1,0.86)
DF11–23	(0.96,0.51,0.51,0.04)	(0.75,0.67,0.87)	0.09	(0.56,0.83,0.87)	0	(0.72,0.95,0.98)
DF11–24	(0.98,0.33,0.33,0.02)	(0.75,0.09,0.57)	0.04	(0.67,1,0.79)	0	(0.63,1.02,0.5)
DF15–4	(0.94,0.23,0.23,0.06)	(0.5,2.57,1.86)	0.12	(0.75,1.14,0.87)	0	(0.63,1.01,0.25)
DF15–5	(0.96,0.32,0.32,0.04)	(0.5,0.09,0.57)	0.08	(0.68,0.96,0.85)	0	(0.64,1.01,0.45)
DF15–20	(0.94,0.42,0.42,0.06)	(0.75,2.57,1.86)	0.12	(0.64,0.99,0.77)	0	(0.71,0.99,0.69)
DF15–21	(0.97,0.37,0.37,0.03)	(0.75,0.09,0.57)	0.07	(0.66,1,0.81)	0	(0.66,1.01,0.58)
DF15–33	(0.96,0.48,0.48,0.04)	(0.9,2.57,1.86)	0.07	(0.55,1,0.87)	0	(0.7,1,0.9)
DF15–35	(0.97,0.51,0.51,0.03)	(0.95,2.57,1.86)	0.07	(0.54,0.82,0.87)	0	(0.7,0.91,0.95)
DRFN08–10	(0.97,0.25,0.25,0.03)	(0.75,2,2)	0.07	(0.75,1,1)	0.03	(0.61,1.03,0.34)
DRFN08–11	(0.95,0.33,0.33,0.05)	(0.75,1,1)	0.1	(0.67,1,0.8)	0	(0.66,1.02,0.45)
DO09–32	(0.95,0.39,0.39,0.05)	(0.9,1,1)	0.09	(0.65,1,0.8)	0	(0.68,1,0.62)
FY17–25	(0.96,0.35,0.35,0.04)	(0.75,0.4,0.4)	0.09	(0.66,0.99,0.79)	0	(0.66,1,0.51)
FRD12–29	(0.96,0.54,0.54,0.04)	(0.88,0.33,0.33)	0.07	(0.53,0.82,1)	0	(0.68,0.8,0.93)
KS13–12	(0.96,0.36,0.36,0.04)	(0.75,1,0.5)	0.07	(0.66,0.99,0.81)	0	(0.65,1.01,0.54)
STS13–13	(0.95,0.55,0.55,0.05)	(0.75,1,0.25)	0.09	(0.59,0.61,0.9)	0.01	(0.73,0.88,1.08)
Means			0.088	(0.65,0.95,0.86)	0.007	(0.67,0.98,0.59)

Note: "SG-Strategy" is the SG-continuation strategy estimated in the "1.5× SG + AD" model

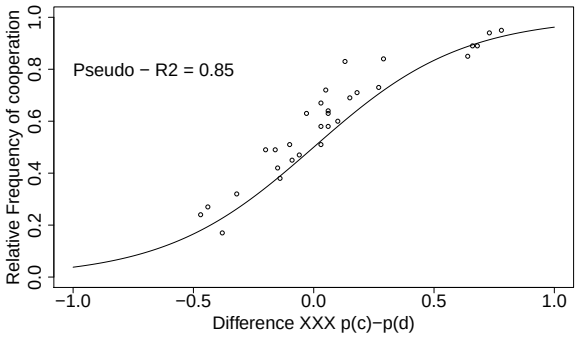
Table 51: Distance to Semi-Grim MPEs: closest equilibria (second halves of sessions)

Treatment	Empirical		Closest Reg-SG MPE	Closest BF-SG MPE
	SG-Strategy	Game	Strategy	Strategy
AF09–34	(0.97,0.46,0.46,0.03)	(0.9,0.33,0.11)	(1,0.46,0.46,0)	(0.97,0.46,0.46,0.03)
BOS11–9	(1,0.13,0.13,0)	(0.5,2,2)	(1,0.13,0.13,0)	(1,0.13,0.13,0)
BOS11–14	(0.99,0.3,0.3,0.01)	(0.75,0.5,3.5)	(1,0.3,0.3,0)	(0.99,0.3,0.3,0.01)
BOS11–15	(1,0,0,0)	(0.75,1,8)	(1,0,0,0)	(0.93,0.08,0.08,0.07)
BOS11–16	(0.97,0.21,0.21,0.03)	(0.75,0.75,1.25)	(1,0.21,0.21,0)	(0.97,0.24,0.24,0.03)
BOS11–17	(0.95,0.26,0.26,0.05)	(0.75,0.83,0.5)	(1,0.26,0.26,0)	(0.96,0.29,0.29,0.04)
BOS11–26	(0.94,0.29,0.29,0.06)	(0.75,2,2)	(1,0.29,0.29,0)	(0.94,0.29,0.29,0.06)
BOS11–27	(0.95,0.5,0.5,0.05)	(0.75,1,1)	(1,0.5,0.5,0)	(0.95,0.5,0.5,0.05)
BOS11–30	(0.96,0.2,0.2,0.04)	(0.88,0.5,3.5)	(1,0.2,0.2,0)	(0.97,0.2,0.2,0.03)
BOS11–31	(0.98,0.48,0.48,0.02)	(0.88,2,2)	(1,0.48,0.48,0)	(0.98,0.48,0.48,0.02)
BK12–28	(0.95,0.32,0.32,0.05)	(0.8,1.17,0.83)	(1,0.32,0.32,0)	(0.95,0.32,0.32,0.05)
D05–18	(0.88,0.4,0.4,0.12)	(0.75,1.17,0.83)	(1,0.4,0.4,0)	(0.88,0.4,0.4,0.12)
D05–19	(0.95,0.3,0.3,0.05)	(0.75,0.83,1.17)	(1,0.3,0.3,0)	(0.95,0.3,0.3,0.05)
DF11–6	(0.94,0.55,0.55,0.06)	(0.5,2.57,1.86)	(1,0.55,0.55,0)	(0.94,0.55,0.55,0.06)
DF11–7	(0.86,0.47,0.47,0.14)	(0.5,0.67,0.87)	(1,0.47,0.47,0)	(0.86,0.47,0.47,0.14)
DF11–8	(0.97,0.45,0.45,0.03)	(0.5,0.09,0.57)	(1,0.45,0.45,0)	(0.97,0.45,0.45,0.03)
DF11–22	(0.96,0.47,0.47,0.04)	(0.75,2.57,1.86)	(1,0.47,0.47,0)	(0.96,0.47,0.47,0.04)
DF11–23	(0.96,0.51,0.51,0.04)	(0.75,0.67,0.87)	(1,0.51,0.51,0)	(0.96,0.51,0.51,0.04)
DF11–24	(0.98,0.33,0.33,0.02)	(0.75,0.09,0.57)	(1,0.33,0.33,0)	(0.98,0.33,0.33,0.02)
DF15–4	(0.94,0.23,0.23,0.06)	(0.5,2.57,1.86)	(1,0.23,0.23,0)	(0.94,0.23,0.23,0.06)
DF15–5	(0.96,0.32,0.32,0.04)	(0.5,0.09,0.57)	(1,0.32,0.32,0)	(0.96,0.32,0.32,0.04)
DF15–20	(0.94,0.42,0.42,0.06)	(0.75,2.57,1.86)	(1,0.42,0.42,0)	(0.94,0.42,0.42,0.06)
DF15–21	(0.97,0.37,0.37,0.03)	(0.75,0.09,0.57)	(1,0.37,0.37,0)	(0.97,0.37,0.37,0.03)
DF15–33	(0.96,0.48,0.48,0.04)	(0.9,2.57,1.86)	(1,0.48,0.48,0)	(0.96,0.48,0.48,0.04)
DF15–35	(0.97,0.51,0.51,0.03)	(0.95,2.57,1.86)	(1,0.51,0.51,0)	(0.97,0.51,0.51,0.03)
DRFN08–10	(0.97,0.25,0.25,0.03)	(0.75,2,2)	(1,0.25,0.25,0)	(0.98,0.26,0.26,0.02)
DRFN08–11	(0.95,0.33,0.33,0.05)	(0.75,1,1)	(1,0.33,0.33,0)	(0.95,0.33,0.33,0.05)
DO09–32	(0.95,0.39,0.39,0.05)	(0.9,1,1)	(1,0.39,0.39,0)	(0.95,0.39,0.39,0.05)
FY17–25	(0.96,0.35,0.35,0.04)	(0.75,0.4,0.4)	(1,0.35,0.35,0)	(0.96,0.35,0.35,0.04)
FRD12–29	(0.96,0.54,0.54,0.04)	(0.88,0.33,0.33)	(1,0.54,0.54,0)	(0.96,0.54,0.54,0.04)
KS13–12	(0.96,0.36,0.36,0.04)	(0.75,1,0.5)	(1,0.36,0.36,0)	(0.96,0.36,0.36,0.04)
STS13–13	(0.95,0.55,0.55,0.05)	(0.75,1,0.25)	(1,0.55,0.55,0)	(0.95,0.55,0.55,0.05)

Note: "SG-Strategy" is the SG-continuation strategy estimated in the "1.5× SG + AD" model

Table 52: Incentives in state $\emptyset$ (second halves of sessions)

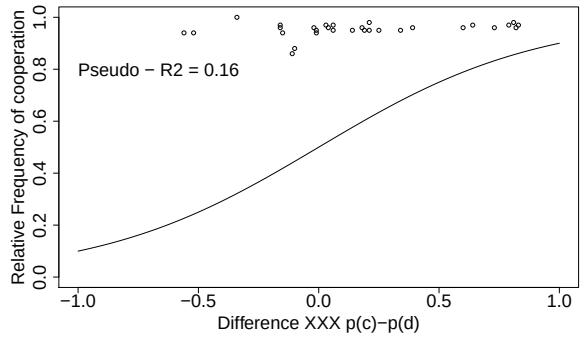
Treatment	Game	Observation			Fit	
		$\hat{\sigma}_0$	$\hat{\pi}(c)$	$\hat{\pi}(d)$	σ_0^*	Deviat
AF09–34	(0.9,0.33,0.11)	0.94	7.7	6.46	0.99	-0.05
BOS11–9	(0.5,2,2)	0.27	0.75	1.18	0.19	0.08
BOS11–14	(0.75,0.5,3.5)	0.03	0.75	1	0.3	-0.27
BOS11–15	(0.75,1,8)	0	0.75	1	0.3	-0.3
BOS11–16	(0.75,0.75,1.25)	0.63	1.32	1.28	0.53	0.1
BOS11–17	(0.75,0.83,0.5)	0.6	1.69	1.5	0.66	-0.06
BOS11–26	(0.75,2,2)	0.49	0.99	1.12	0.39	0.1
BOS11–27	(0.75,1,1)	0.47	1.27	1.23	0.53	-0.06
BOS11–30	(0.88,0.5,3.5)	0.45	0.95	1	0.46	-0.01
BOS11–31	(0.88,2,2)	0.58	1.11	1.09	0.52	0.06
BK12–28	(0.8,1.17,0.83)	0.58	1.44	1.36	0.57	0.01
D05–18	(0.75,1.17,0.83)	0.51	1.5	1.59	0.42	0.09
D05–19	(0.75,0.83,1.17)	0.67	1.3	1.26	0.53	0.14
DF11–6	(0.5,2.57,1.86)	0.17	0.76	1.07	0.26	-0.09
DF11–7	(0.5,0.67,0.87)	0.32	0.98	1.24	0.29	0.03
DF11–8	(0.5,0.09,0.57)	0.64	1.53	1.43	0.58	0.06
DF11–22	(0.75,2.57,1.86)	0.38	1.08	1.18	0.42	-0.04
DF11–23	(0.75,0.67,0.87)	0.83	1.79	1.61	0.65	0.18
DF11–24	(0.75,0.09,0.57)	0.95	2.54	1.73	0.94	0.01
DF15–4	(0.5,2.57,1.86)	0.24	0.68	1.1	0.19	0.05
DF15–5	(0.5,0.09,0.57)	0.69	1.65	1.52	0.61	0.08
DF15–20	(0.75,2.57,1.86)	0.42	1.06	1.17	0.41	0.01
DF15–21	(0.75,0.09,0.57)	0.85	2.19	1.56	0.89	-0.04
DF15–33	(0.9,2.57,1.86)	0.51	1.22	1.18	0.53	-0.02
DF15–35	(0.95,2.57,1.86)	0.63	1.27	1.21	0.55	0.08
DRFN08–10	(0.75,2,2)	0.49	0.98	1.11	0.39	0.1
DRFN08–11	(0.75,1,1)	0.72	1.52	1.44	0.57	0.15
DO09–32	(0.9,1,1)	0.71	1.53	1.36	0.64	0.07
FY17–25	(0.75,0.4,0.4)	0.89	2.66	1.95	0.92	-0.03
FRD12–29	(0.88,0.33,0.33)	0.89	3.18	2.41	0.93	-0.04
KS13–12	(0.75,1,0.5)	0.84	2.29	1.93	0.77	0.07
STS13–13	(0.75,1,0.25)	0.73	3.87	3.39	0.84	-0.11



Note: For each treatment in each experiment, the table reviews the treatment parameters, the observed relative frequency of cooperation (in state $\emptyset$ in second halves of sessions), the expected payoff cooperating in that state $\hat{\pi}(c)$, the expected payoff of defecting in that state $\hat{\pi}(d)$, the "predicted" probability of cooperation based on the logistic regression of cooperation rates on monetary incentive $\hat{\pi}(c) - \hat{\pi}(c)$, and the absolute deviation of that prediction.

Table 53: Incentives in state *cc* (second halves of sessions)

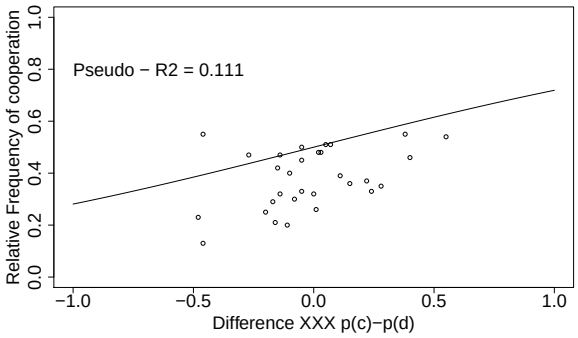
Treatment	Game	Observation			Fit	
		$\hat{\sigma}_{cc}$	$\hat{\pi}(c)$	$\hat{\pi}(d)$	σ_0^*	Deviat
AF09–34	(0.9,0.33,0.11)	0.97	7.86	6.61	0.98	-0.01
BOS11–9	(0.5,2,2)	1	1.46	1.75	0.28	0.72
BOS11–14	(0.75,0.5,3.5)	0.99	1.25	1.1	0.62	0.37
BOS11–15	(0.75,1,8)	1	1.12	1.06	0.55	0.45
BOS11–16	(0.75,0.75,1.25)	0.97	1.59	1.42	0.63	0.34
BOS11–17	(0.75,0.83,0.5)	0.95	2.57	2.03	0.85	0.1
BOS11–26	(0.75,2,2)	0.94	1.35	1.37	0.48	0.46
BOS11–27	(0.75,1,1)	0.95	1.8	1.62	0.64	0.31
BOS11–30	(0.88,0.5,3.5)	0.96	1.15	1.03	0.59	0.37
BOS11–31	(0.88,2,2)	0.98	1.38	1.23	0.62	0.36
BK12–28	(0.8,1.17,0.83)	0.95	1.9	1.65	0.69	0.26
D05–18	(0.75,1.17,0.83)	0.88	1.83	1.9	0.44	0.44
D05–19	(0.75,0.83,1.17)	0.95	1.64	1.43	0.66	0.29
DF11–6	(0.5,2.57,1.86)	0.94	1.42	1.85	0.2	0.74
DF11–7	(0.5,0.67,0.87)	0.86	1.81	1.9	0.43	0.43
DF11–8	(0.5,0.09,0.57)	0.97	2.61	2	0.88	0.09
DF11–22	(0.75,2.57,1.86)	0.96	1.46	1.56	0.42	0.54
DF11–23	(0.75,0.67,0.87)	0.96	1.96	1.74	0.67	0.29
DF11–24	(0.75,0.09,0.57)	0.98	2.59	1.75	0.94	0.04
DF15–4	(0.5,2.57,1.86)	0.94	1.41	1.87	0.19	0.75
DF15–5	(0.5,0.09,0.57)	0.96	2.57	1.97	0.87	0.09
DF15–20	(0.75,2.57,1.86)	0.94	1.42	1.52	0.42	0.52
DF15–21	(0.75,0.09,0.57)	0.97	2.47	1.68	0.93	0.04
DF15–33	(0.9,2.57,1.86)	0.96	1.4	1.34	0.55	0.41
DF15–35	(0.95,2.57,1.86)	0.97	1.36	1.29	0.56	0.41
DRFN08–10	(0.75,2,2)	0.97	1.4	1.37	0.52	0.45
DRFN08–11	(0.75,1,1)	0.95	1.79	1.62	0.63	0.32
DO09–32	(0.9,1,1)	0.95	1.67	1.45	0.67	0.28
FY17–25	(0.75,0.4,0.4)	0.96	3.03	2.15	0.94	0.02
FRD12–29	(0.88,0.33,0.33)	0.96	3.38	2.57	0.93	0.03
KS13–12	(0.75,1,0.5)	0.96	2.71	2.25	0.81	0.15
STS13–13	(0.75,1,0.25)	0.95	4.54	4.32	0.67	0.28



Note: For each treatment in each experiment, the table reviews the treatment parameters, the observed relative frequency of cooperation (in state *cc* in second halves of sessions), the expected payoff cooperating in that state $\hat{\pi}(c)$, the expected payoff of defecting in that state $\hat{\pi}(d)$, the "predicted" probability of cooperation based on the logistic regression of cooperation rates on monetary incentive $\hat{\pi}(c) - \hat{\pi}(c)$, and the absolute deviation of that prediction.

Table 54: Incentives in state cd, dc (second halves of sessions)

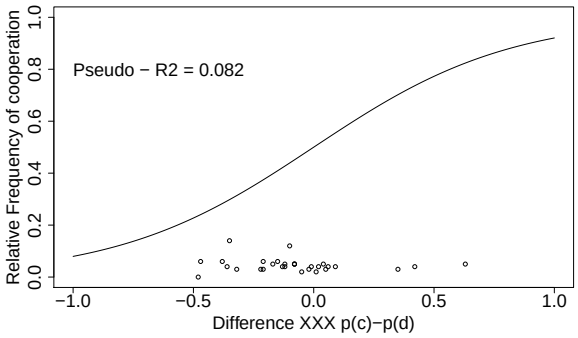
Treatment	Game	Observation			Fit	
		$\hat{\sigma}_{cd,dc}$	$\hat{\pi}(c)$	$\hat{\pi}(d)$	σ_0^*	Deviat
AF09–34	(0.9,0.33,0.11)	0.46	6.66	5.49	0.63	-0.17
BOS11–9	(0.5,2,2)	0.13	0.64	1.1	0.45	-0.32
BOS11–14	(0.75,0.5,3.5)	0.3	0.9	1.02	0.49	-0.19
BOS11–15	(0.75,1,8)	0	0.75	1	0.47	-0.47
BOS11–16	(0.75,0.75,1.25)	0.21	1.06	1.15	0.49	-0.28
BOS11–17	(0.75,0.83,0.5)	0.26	1.42	1.33	0.51	-0.25
BOS11–26	(0.75,2,2)	0.29	0.99	1.12	0.48	-0.19
BOS11–27	(0.75,1,1)	0.5	1.41	1.33	0.51	-0.01
BOS11–30	(0.88,0.5,3.5)	0.2	0.92	1	0.49	-0.29
BOS11–31	(0.88,2,2)	0.48	1.15	1.11	0.5	-0.02
BK12–28	(0.8,1.17,0.83)	0.32	1.31	1.27	0.5	-0.18
D05–18	(0.75,1.17,0.83)	0.4	1.44	1.52	0.49	-0.09
D05–19	(0.75,0.83,1.17)	0.3	1.09	1.15	0.49	-0.19
DF11–6	(0.5,2.57,1.86)	0.55	1.11	1.49	0.46	0.09
DF11–7	(0.5,0.67,0.87)	0.47	1.31	1.51	0.48	-0.01
DF11–8	(0.5,0.09,0.57)	0.45	1.57	1.45	0.51	-0.06
DF11–22	(0.75,2.57,1.86)	0.47	1.19	1.28	0.49	-0.02
DF11–23	(0.75,0.67,0.87)	0.51	1.56	1.43	0.52	-0.01
DF11–24	(0.75,0.09,0.57)	0.33	1.57	1.29	0.53	-0.2
DF15–4	(0.5,2.57,1.86)	0.23	0.79	1.22	0.45	-0.22
DF15–5	(0.5,0.09,0.57)	0.32	1.26	1.32	0.49	-0.17
DF15–20	(0.75,2.57,1.86)	0.42	1.13	1.24	0.49	-0.07
DF15–21	(0.75,0.09,0.57)	0.37	1.58	1.29	0.53	-0.16
DF15–33	(0.9,2.57,1.86)	0.48	1.24	1.2	0.5	-0.02
DF15–35	(0.95,2.57,1.86)	0.51	1.27	1.21	0.51	0
DRFN08–10	(0.75,2,2)	0.25	0.95	1.1	0.48	-0.23
DRFN08–11	(0.75,1,1)	0.33	1.23	1.25	0.5	-0.17
DO09–32	(0.9,1,1)	0.39	1.38	1.27	0.51	-0.12
FY17–25	(0.75,0.4,0.4)	0.35	1.83	1.49	0.54	-0.19
FRD12–29	(0.88,0.33,0.33)	0.54	2.74	2.06	0.58	-0.04
KS13–12	(0.75,1,0.5)	0.36	1.73	1.51	0.53	-0.17
STS13–13	(0.75,1,0.25)	0.55	3.64	3.07	0.57	-0.02



Note: For each treatment in each experiment, the table reviews the treatment parameters, the observed relative frequency of cooperation (in states cd, dc in second halves of sessions), the expected payoff cooperating in that state $\hat{\pi}(c)$, the expected payoff of defecting in that state $\hat{\pi}(d)$, the "predicted" probability of cooperation based on the logistic regression of cooperation rates on monetary incentive $\hat{\pi}(c) - \hat{\pi}(c)$, and the absolute deviation of that prediction.

Table 55: Incentives in state *dd* (second halves of sessions)

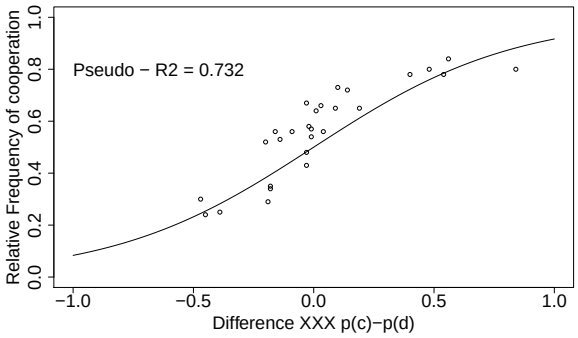
Treatment	Game	Observation			Fit	
		$\hat{\sigma}_{dd}$	$\hat{\pi}(c)$	$\hat{\pi}(d)$	σ_0^*	Deviat
AF09–34	(0.9,0.33,0.11)	0.03	5.71	4.6	0.64	-0.61
BOS11–9	(0.5,2,2)	0	0.54	1.02	0.44	-0.44
BOS11–14	(0.75,0.5,3.5)	0.01	0.75	1	0.47	-0.46
BOS11–15	(0.75,1,8)	0	0.75	1	0.47	-0.47
BOS11–16	(0.75,0.75,1.25)	0.03	0.91	1.08	0.48	-0.45
BOS11–17	(0.75,0.83,0.5)	0.05	1.06	1.12	0.49	-0.44
BOS11–26	(0.75,2,2)	0.06	0.85	1.03	0.48	-0.42
BOS11–27	(0.75,1,1)	0.05	1.04	1.06	0.5	-0.45
BOS11–30	(0.88,0.5,3.5)	0.04	0.87	0.99	0.48	-0.44
BOS11–31	(0.88,2,2)	0.02	0.99	1.03	0.49	-0.47
BK12–28	(0.8,1.17,0.83)	0.05	1.06	1.11	0.49	-0.44
D05–18	(0.75,1.17,0.83)	0.12	1.21	1.3	0.49	-0.37
D05–19	(0.75,0.83,1.17)	0.05	0.89	1.05	0.48	-0.43
DF11–6	(0.5,2.57,1.86)	0.06	0.73	1.03	0.46	-0.4
DF11–7	(0.5,0.67,0.87)	0.14	0.82	1.12	0.46	-0.32
DF11–8	(0.5,0.09,0.57)	0.03	0.77	1.03	0.47	-0.44
DF11–22	(0.75,2.57,1.86)	0.04	0.95	1.05	0.49	-0.45
DF11–23	(0.75,0.67,0.87)	0.04	1.14	1.1	0.51	-0.47
DF11–24	(0.75,0.09,0.57)	0.02	1.07	1.06	0.5	-0.48
DF15–4	(0.5,2.57,1.86)	0.06	0.61	1.04	0.44	-0.38
DF15–5	(0.5,0.09,0.57)	0.04	0.71	1.05	0.46	-0.42
DF15–20	(0.75,2.57,1.86)	0.06	0.93	1.05	0.48	-0.42
DF15–21	(0.75,0.09,0.57)	0.03	1.07	1.07	0.5	-0.47
DF15–33	(0.9,2.57,1.86)	0.04	1.1	1.08	0.5	-0.46
DF15–35	(0.95,2.57,1.86)	0.03	1.17	1.12	0.51	-0.48
DRFN08–10	(0.75,2,2)	0.03	0.82	1.02	0.47	-0.44
DRFN08–11	(0.75,1,1)	0.05	0.97	1.08	0.49	-0.44
DO09–32	(0.9,1,1)	0.05	1.2	1.16	0.51	-0.46
FY17–25	(0.75,0.4,0.4)	0.04	1.24	1.16	0.51	-0.47
FRD12–29	(0.88,0.33,0.33)	0.04	2.1	1.55	0.57	-0.53
KS13–12	(0.75,1,0.5)	0.04	1.2	1.12	0.51	-0.47
STS13–13	(0.75,1,0.25)	0.05	2.54	1.55	0.63	-0.58



Note: For each treatment in each experiment, the table reviews the treatment parameters, the observed relative frequency of cooperation (in state *dd* in second halves of sessions), the expected payoff cooperating in that state $\hat{\pi}(c)$, the expected payoff of defecting in that state $\hat{\pi}(d)$, the "predicted" probability of cooperation based on the logistic regression of cooperation rates on monetary incentive $\hat{\pi}(c) - \hat{\pi}(c)$, and the absolute deviation of that prediction.

Table 56: Incentives in state 0 (first halves of sessions)

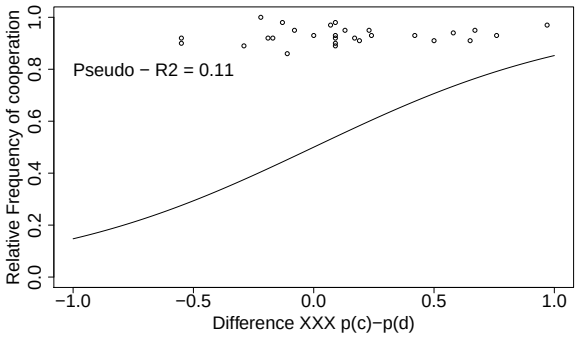
Treatment	Game	Observation			Fit	
		$\hat{\sigma}_\emptyset$	$\hat{\pi}(c)$	$\hat{\pi}(d)$	σ_0^*	Deviat
AF09–34	(0.9,0.33,0.11)	0.78	6.24	5.51	0.87	-0.09
BOS11–9	(0.5,2,2)	0.36	0.76	1.18	0.26	0.1
BOS11–14	(0.75,0.5,3.5)	0.11	0.78	0.99	0.37	-0.26
BOS11–15	(0.75,1,8)	0.2	0.76	1	0.35	-0.15
BOS11–16	(0.75,0.75,1.25)	0.57	1.26	1.25	0.51	0.06
BOS11–17	(0.75,0.83,0.5)	0.52	1.7	1.83	0.42	0.1
BOS11–26	(0.75,2,2)	0.29	0.96	1.14	0.39	-0.1
BOS11–27	(0.75,1,1)	0.56	1.17	1.2	0.48	0.08
BOS11–30	(0.88,0.5,3.5)	0.69	0.95	0.99	0.47	0.22
BOS11–31	(0.88,2,2)	0.64	1.18	1.14	0.53	0.11
BK12–28	(0.8,1.17,0.83)	0.54	1.43	1.44	0.49	0.05
D05–18	(0.75,1.17,0.83)	0.53	1.43	1.54	0.43	0.1
D05–19	(0.75,0.83,1.17)	0.58	1.21	1.24	0.48	0.1
DF11–6	(0.5,2.57,1.86)	0.24	0.77	1.15	0.27	-0.03
DF11–7	(0.5,0.67,0.87)	0.25	0.89	1.22	0.3	-0.05
DF11–8	(0.5,0.09,0.57)	0.48	1.41	1.43	0.49	-0.01
DF11–22	(0.75,2.57,1.86)	0.35	1.04	1.19	0.41	-0.06
DF11–23	(0.75,0.67,0.87)	0.65	1.51	1.42	0.56	0.09
DF11–24	(0.75,0.09,0.57)	0.8	2.04	1.6	0.75	0.05
DF15–4	(0.5,2.57,1.86)	0.3	0.75	1.16	0.26	0.04
DF15–5	(0.5,0.09,0.57)	0.73	1.68	1.59	0.56	0.17
DF15–20	(0.75,2.57,1.86)	0.34	1.02	1.18	0.4	-0.06
DF15–21	(0.75,0.09,0.57)	0.78	1.94	1.56	0.73	0.05
DF15–33	(0.9,2.57,1.86)	0.43	1.13	1.16	0.48	-0.05
DF15–35	(0.95,2.57,1.86)	0.56	1.24	1.21	0.52	0.04
DRFN08–10	(0.75,2,2)	0.56	1.08	1.2	0.42	0.14
DRFN08–11	(0.75,1,1)	0.67	1.34	1.35	0.49	0.18
DO09–32	(0.9,1,1)	0.66	1.35	1.32	0.52	0.14
FY17–25	(0.75,0.4,0.4)	0.84	2.43	1.87	0.81	0.03
FRD12–29	(0.88,0.33,0.33)	0.8	3.06	2.18	0.9	-0.1
KS13–12	(0.75,1,0.5)	0.72	1.99	1.81	0.61	0.11
STS13–13	(0.75,1,0.25)	0.65	3.33	3.09	0.65	0



Note: For each treatment in each experiment, the table reviews the treatment parameters, the observed relative frequency of cooperation (in state $\emptyset$ in first halves of sessions), the expected payoff cooperating in that state $\hat{\pi}(c)$, the expected payoff of defecting in that state $\hat{\pi}(d)$, the "predicted" probability of cooperation based on the logistic regression of cooperation rates on monetary incentive $\hat{\pi}(c) - \hat{\pi}(c)$, and the absolute deviation of that prediction.

Table 57: Incentives in state *cc* (first halves of sessions)

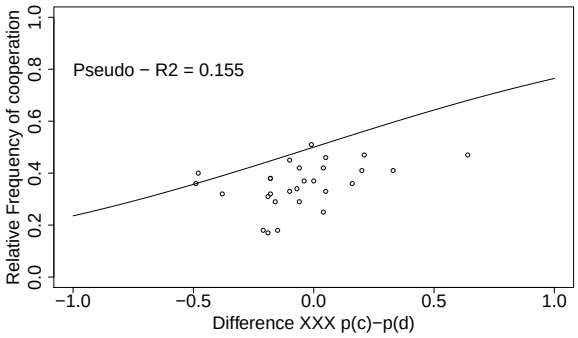
Treatment	Game	Observation			Fit	
		$\hat{\sigma}_{cc}$	$\hat{\pi}(c)$	$\hat{\pi}(d)$	σ_0^*	Deviat
AF09–34	(0.9,0.33,0.11)	0.91	6.53	5.77	0.85	0.06
BOS11–9	(0.5,2,2)	0.95	1.41	1.71	0.33	0.62
BOS11–14	(0.75,0.5,3.5)	0.99	1.09	1.07	0.51	0.48
BOS11–15	(0.75,1,8)	1	1.05	1.03	0.51	0.49
BOS11–16	(0.75,0.75,1.25)	0.95	1.55	1.39	0.59	0.36
BOS11–17	(0.75,0.83,0.5)	1	2.08	2.21	0.43	0.57
BOS11–26	(0.75,2,2)	0.98	1.33	1.39	0.47	0.51
BOS11–27	(0.75,1,1)	0.89	1.6	1.52	0.55	0.34
BOS11–30	(0.88,0.5,3.5)	1	1.2	1.03	0.6	0.4
BOS11–31	(0.88,2,2)	0.98	1.39	1.26	0.57	0.41
BK12–28	(0.8,1.17,0.83)	0.92	1.79	1.7	0.55	0.37
D05–18	(0.75,1.17,0.83)	0.86	1.74	1.8	0.47	0.39
D05–19	(0.75,0.83,1.17)	0.91	1.56	1.43	0.57	0.34
DF11–6	(0.5,2.57,1.86)	0.92	1.4	1.88	0.25	0.67
DF11–7	(0.5,0.67,0.87)	0.89	1.84	1.92	0.45	0.44
DF11–8	(0.5,0.09,0.57)	0.91	2.38	2	0.71	0.2
DF11–22	(0.75,2.57,1.86)	0.92	1.39	1.53	0.42	0.5
DF11–23	(0.75,0.67,0.87)	0.95	1.89	1.67	0.62	0.33
DF11–24	(0.75,0.09,0.57)	0.95	2.32	1.73	0.8	0.15
DF15–4	(0.5,2.57,1.86)	0.9	1.33	1.81	0.25	0.65
DF15–5	(0.5,0.09,0.57)	0.93	2.36	1.95	0.72	0.21
DF15–20	(0.75,2.57,1.86)	0.92	1.39	1.52	0.43	0.49
DF15–21	(0.75,0.09,0.57)	0.94	2.3	1.75	0.78	0.16
DF15–33	(0.9,2.57,1.86)	0.93	1.31	1.29	0.51	0.42
DF15–35	(0.95,2.57,1.86)	0.97	1.33	1.28	0.53	0.44
DRFN08–10	(0.75,2,2)	0.95	1.37	1.38	0.49	0.46
DRFN08–11	(0.75,1,1)	0.93	1.69	1.57	0.57	0.36
DO09–32	(0.9,1,1)	0.9	1.48	1.4	0.55	0.35
FY17–25	(0.75,0.4,0.4)	0.93	2.78	2.06	0.84	0.09
FRD12–29	(0.88,0.33,0.33)	0.97	3.4	2.43	0.9	0.07
KS13–12	(0.75,1,0.5)	0.93	2.52	2.23	0.66	0.27
STS13–13	(0.75,1,0.25)	0.92	4.13	3.96	0.6	0.32



Note: For each treatment in each experiment, the table reviews the treatment parameters, the observed relative frequency of cooperation (in state *cc* in first halves of sessions), the expected payoff cooperating in that state $\hat{\pi}(c)$, the expected payoff of defecting in that state $\hat{\pi}(d)$, the "predicted" probability of cooperation based on the logistic regression of cooperation rates on monetary incentive $\hat{\pi}(c) - \hat{\pi}(c)$, and the absolute deviation of that prediction.

Table 58: Incentives in state cd, dc (first halves of sessions)

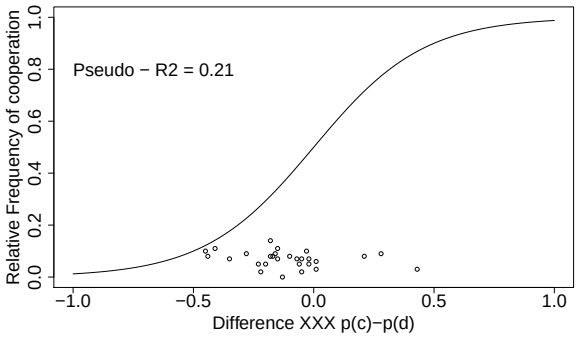
Treatment	Game	Observation			Fit	
		$\hat{\sigma}_{cd,dc}$	$\hat{\pi}(c)$	$\hat{\pi}(d)$	σ_0^*	Deviat
AF09–34	(0.9,0.33,0.11)	0.41	5.62	4.94	0.69	-0.28
BOS11–9	(0.5,2,2)	0.2	0.73	1.15	0.38	-0.18
BOS11–14	(0.75,0.5,3.5)	0.12	0.8	1	0.44	-0.32
BOS11–15	(0.75,1,8)	0.22	0.79	1	0.44	-0.22
BOS11–16	(0.75,0.75,1.25)	0.18	0.99	1.12	0.46	-0.28
BOS11–17	(0.75,0.83,0.5)	0.38	1.63	1.76	0.46	-0.08
BOS11–26	(0.75,2,2)	0.17	0.91	1.1	0.44	-0.27
BOS11–27	(0.75,1,1)	0.45	1.27	1.28	0.5	-0.05
BOS11–30	(0.88,0.5,3.5)	0	0.88	0.98	0.47	-0.47
BOS11–31	(0.88,2,2)	0.51	1.18	1.14	0.51	0
BK12–28	(0.8,1.17,0.83)	0.29	1.27	1.32	0.49	-0.2
D05–18	(0.75,1.17,0.83)	0.29	1.26	1.4	0.46	-0.17
D05–19	(0.75,0.83,1.17)	0.34	1.13	1.19	0.48	-0.14
DF11–6	(0.5,2.57,1.86)	0.4	0.97	1.38	0.38	0.02
DF11–7	(0.5,0.67,0.87)	0.32	1.07	1.36	0.42	-0.1
DF11–8	(0.5,0.09,0.57)	0.42	1.5	1.49	0.5	-0.08
DF11–22	(0.75,2.57,1.86)	0.38	1.11	1.25	0.46	-0.08
DF11–23	(0.75,0.67,0.87)	0.46	1.41	1.35	0.52	-0.06
DF11–24	(0.75,0.09,0.57)	0.36	1.49	1.35	0.54	-0.18
DF15–4	(0.5,2.57,1.86)	0.36	0.89	1.32	0.38	-0.02
DF15–5	(0.5,0.09,0.57)	0.31	1.15	1.32	0.45	-0.14
DF15–20	(0.75,2.57,1.86)	0.32	1.06	1.21	0.46	-0.14
DF15–21	(0.75,0.09,0.57)	0.47	1.65	1.42	0.57	-0.1
DF15–33	(0.9,2.57,1.86)	0.37	1.14	1.16	0.49	-0.12
DF15–35	(0.95,2.57,1.86)	0.42	1.2	1.19	0.5	-0.08
DRFN08–10	(0.75,2,2)	0.18	0.89	1.09	0.44	-0.26
DRFN08–11	(0.75,1,1)	0.33	1.16	1.23	0.48	-0.15
DO09–32	(0.9,1,1)	0.37	1.27	1.27	0.5	-0.13
FY17–25	(0.75,0.4,0.4)	0.25	1.51	1.39	0.53	-0.28
FRD12–29	(0.88,0.33,0.33)	0.47	2.6	1.85	0.71	-0.24
KS13–12	(0.75,1,0.5)	0.33	1.64	1.54	0.53	-0.2
STS13–13	(0.75,1,0.25)	0.41	2.93	2.64	0.58	-0.17



Note: For each treatment in each experiment, the table reviews the treatment parameters, the observed relative frequency of cooperation (in state cd, dc in first halves of sessions), the expected payoff cooperating in that state $\hat{\pi}(c)$, the expected payoff of defecting in that state $\hat{\pi}(d)$, the "predicted" probability of cooperation based on the logistic regression of cooperation rates on monetary incentive $\hat{\pi}(c) - \hat{\pi}(c)$, and the absolute deviation of that prediction.

Table 59: Incentives in state *dd* (first halves of sessions)

Treatment	Game	Observation			Fit	
		$\hat{\sigma}_{dd}$	$\hat{\pi}(c)$	$\hat{\pi}(d)$	σ_0^*	Deviat
AF09–34	(0.9,0.33,0.11)	0.09	5	4.39	0.87	-0.78
BOS11–9	(0.5,2,2)	0.05	0.58	1.03	0.2	-0.15
BOS11–14	(0.75,0.5,3.5)	0.01	0.74	0.99	0.32	-0.31
BOS11–15	(0.75,1,8)	0	0.75	1	0.32	-0.32
BOS11–16	(0.75,0.75,1.25)	0.05	0.87	1.06	0.36	-0.31
BOS11–17	(0.75,0.83,0.5)	0	1.46	1.59	0.4	-0.4
BOS11–26	(0.75,2,2)	0.02	0.83	1.04	0.34	-0.32
BOS11–27	(0.75,1,1)	0.11	0.98	1.07	0.43	-0.32
BOS11–30	(0.88,0.5,3.5)	0	0.87	0.98	0.42	-0.42
BOS11–31	(0.88,2,2)	0.02	1.02	1.04	0.48	-0.46
BK12–28	(0.8,1.17,0.83)	0.08	1.11	1.21	0.42	-0.34
D05–18	(0.75,1.17,0.83)	0.14	1.13	1.29	0.38	-0.24
D05–19	(0.75,0.83,1.17)	0.09	0.92	1.07	0.39	-0.3
DF11–6	(0.5,2.57,1.86)	0.08	0.69	1.06	0.24	-0.16
DF11–7	(0.5,0.67,0.87)	0.11	0.75	1.12	0.24	-0.13
DF11–8	(0.5,0.09,0.57)	0.09	0.85	1.11	0.31	-0.22
DF11–22	(0.75,2.57,1.86)	0.08	0.94	1.09	0.39	-0.31
DF11–23	(0.75,0.67,0.87)	0.05	1.08	1.14	0.45	-0.4
DF11–24	(0.75,0.09,0.57)	0.05	1.16	1.2	0.47	-0.42
DF15–4	(0.5,2.57,1.86)	0.1	0.67	1.07	0.22	-0.12
DF15–5	(0.5,0.09,0.57)	0.07	0.8	1.14	0.26	-0.19
DF15–20	(0.75,2.57,1.86)	0.08	0.91	1.08	0.37	-0.29
DF15–21	(0.75,0.09,0.57)	0.06	1.21	1.19	0.52	-0.46
DF15–33	(0.9,2.57,1.86)	0.07	1.06	1.1	0.47	-0.4
DF15–35	(0.95,2.57,1.86)	0.03	1.15	1.16	0.49	-0.46
DRFN08–10	(0.75,2,2)	0.05	0.82	1.04	0.34	-0.29
DRFN08–11	(0.75,1,1)	0.07	0.97	1.11	0.39	-0.32
DO09–32	(0.9,1,1)	0.1	1.18	1.21	0.48	-0.38
FY17–25	(0.75,0.4,0.4)	0.07	1.16	1.21	0.46	-0.39
FRD12–29	(0.88,0.33,0.33)	0.03	1.92	1.37	0.85	-0.82
KS13–12	(0.75,1,0.5)	0.07	1.24	1.22	0.52	-0.45
STS13–13	(0.75,1,0.25)	0.08	2.23	1.88	0.75	-0.67



Note: For each treatment in each experiment, the table reviews the treatment parameters, the observed relative frequency of cooperation (in state *dd* in first halves of sessions), the expected payoff cooperating in that state $\hat{\pi}(c)$, the expected payoff of defecting in that state $\hat{\pi}(d)$, the "predicted" probability of cooperation based on the logistic regression of cooperation rates on monetary incentive $\hat{\pi}(c) - \hat{\pi}(c)$, and the absolute deviation of that prediction.