Backhaus, Teresa; Breitmoser, Yves

Working Paper

God Does Not Play Dice, but Do We? On the Determinism of Choice in Long-run Interactions

Discussion Paper, No. 96

Provided in Cooperation with:
University of Munich (LMU) and Humboldt University Berlin, Collaborative Research Center Transregio 190: Rationality and Competition

Suggested Citation: Backhaus, Teresa; Breitmoser, Yves (2018) : God Does Not Play Dice, but Do We? On the Determinism of Choice in Long-run Interactions, Discussion Paper, No. 96, Ludwig-Maximilians-Universität München und Humboldt-Universität zu Berlin, Collaborative Research Center Transregio 190 - Rationality and Competition, München und Berlin

This Version is available at:
https://hdl.handle.net/10419/185766

God Does Not Play Dice, but Do We?

On the Determinism of Choice in Long-Run Interactions

Teresa Backhaus (WZB)
Yves Breitmoser (HU Berlin)

Discussion Paper No. 96

May 11, 2018

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

God does not play dice, but do we?
On the determinism of choice in long-run interactions

Teresa Backhaus*
BDPEMS and WZB

Yves Breitmoser
Humboldt University Berlin

May 8, 2018

Abstract

When do we cooperate and why? This question concerns one of the most persistent divides between "theory and practice", between predictions from game theory and results from experimental studies. For about 15 years, theoretical analyses predict completely-mixed "behavior" strategies, i.e. strategic randomization rendering "when" and "why" questions largely moot, while experimental analyses seem to consistently identify pure strategies, suggesting long-run interactions are deterministic. Reanalyzing 145,000 decisions from infinitely repeated prisoner's dilemma experiments, and using data-mining techniques giving pure strategies the best possible chance, we conclude that subjects play semi-grim behavior strategies similar to those predicted by theory.

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, Friedel Bolle, Steffen Huck, Johannes Leutgeb, Vlada Pleshcheva, Sebastian Schweighofer-Kodritsch, Roland Strausz, Robert Stüber, Georg Weizsäcker, Roel van Veldhuizen, and Joachim Winter, as well as audiences in Schwanenwerder and Berlin for many helpful comments. Financial support by the Deutsche Forschungsgemeinschaft (BR 4648/1 and CRC TRR 190) is gratefully acknowledged. Address Backhaus: Wissenschaftszentrum Berlin für Sozialforschung, Reichpietschufer 50, 10785 Berlin, Germany, email: teresa.backhaus@wzb.eu, phone: +49 30 25491 411. Address Breitmoser: Spandauer Str. 1, 10099 Berlin, Germany, email: yves.breitmoser@hu-berlin.de, phone: +49 30 2093 99408.

1 Introduction

One of the most dynamic research fields over the last two decades has been behavioral game theory: the econometric and theoretical analysis of laboratory games to align observed behavior with game-theoretical concepts. How should we think of beliefs, utilities and choice of subjects, and is it possible to explain their decisions as responses to incentives? In some classes of games, most notably auctions, behavior seems to be reasonably consistent with theory after accounting simply 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 reflect mainly logistic errors in choice (Weizsäcker, 2003; Brunner et al., 2011); and in games involving the distribution of monetary benefits, interdependence of preferences 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.

A class of games that arguably experienced less progress in aligning behavior and predictions is the large class of repeated games. Repeated games are the main tool in modeling long-run interactions, in particular to study cooperation and defection, and have been a core object of game-theoretic analyses at least since the Folk Theorem for repeated games with discounting (Fudenberg and Maskin, 1986). Part of an explanation for the slow progress may be that theoretical predictions for repeated games tend to be less specific than for other games. The main reason, however, seems to be independent of that: Behavioral analyses of individual strategies in repeated games typically do not assume subgame perfection, even though this normally represents an indisputable assumption in theoretical and behavioral analyses of dynamic games (with the exception of Fudenberg and Levine, 1993). Relaxing subgame perfection is necessary to justify the widespread intuition that human behavior is directly reciprocal, as captured by tit-for-tat in repeated games. When taking tit-for-tat reciprocity as given, dropping subgame perfection in behavioral analyses of repeated games seems inevitable, which perhaps instills the belief that theoretical results maintaining subgame perfection cannot inform such analyses.

This state of affairs raises a provocative question: Who is right? *Theoretical analyses* studying subgame perfect equilibria, assuming subjects understand non-credible threats and seek robustness to, for example, imperfect monitoring (following Kandori, 2002, and Ely and Välimäki, 2002), or *behavioral analyses* analyzing rather complex deterministic choice rules such as tit-for-tat, "lenient grim", or "tit-for-2-tats" (Fudenberg et al., 2012) that mostly violate subgame perfection but seek to build on psychological interpretations of behavior?

With this paper, we seek to tackle this question. Our basis for identification is arguably the main prediction of theoretical analyses of repeated games from the last

twenty years: Following Kandori (2002) and Ely and Välimäki (2002), a large body of theoretical work analyzed robustness to imperfect (private) monitoring and concluded that players seeking robustness are likely to play *behavior strategies* (see also Ely et al., 2005, and Mailath and Samuelson, 2006). Private monitoring is imperfect if opponents' actions are not observable but players get imperfect private signals about their opponents' actions. This covers the case that players fail to perfectly observe or remember their opponents' actions, which seems both empirically and behaviorally plausible. The main prediction of this literature is that subjects should play belief-free equilibria (or weakly belief-free equilibria, Kandori, 2011), i.e. they should randomize each round to ensure the indifference of opponents at all points of time—which implies that opponents do not need to be able to perfectly remember previous actions. This prediction diametrically opposes the assumption of deterministic choice rules made in most recent behavioral analyses (the exception, Breitmoser, 2015, is discussed shortly) and thus provides a powerful foundation for answering the question(s) raised above. The difficulty is that an enormous data set and specific econometric tools are required to conclusively discriminate between behavior strategies, where subjects strategically randomize, and pure strategies, where deviations simply are stochastic mistakes—substantially more data and different tools than in the suggestive analysis of Breitmoser (2015).

The data concern has been solved very recently by Dal Bó and Fréchette (2018), who compiled a very large data set comprising 12 previous experiments (with 32 treatments and around 150,000 decisions), but in their analysis they maintain the assumption of deterministic choice rules. This data set finally allows us to tackle the questions raised above. To conclusively do so, we use a technique that belongs to the realm of data mining: We select models from wide ranges of one- and two-memory pure strategies, as well as from a range of different "switching" rules to capture inconsistency at the individual level. These ranges comprise up to 10^{44} models and represent exhaustive mining for the best combination of pure strategies. Only the best such model will be evaluated against the simple three-parametric "semi-grim" behavior strategy[1] previously identified by Breitmoser (2015)—importantly without penalizing the best model for the selection steps that preceded the evaluation stage. This approach, mining for the best-possible model and then evaluating the model as if it had been hypothesized ex-ante, is called data dredging and is frowned upon for obvious reasons. It drastically biases p-values in favor of the mined model, but in our case this approach helps us to ensure that we are giving pure strategies the best possible chance. Still, we find that not even this exhaustively data-mined mixture improves on the simple hypothesis that subjects play behavior strategies with a rather specific structure (semi-grim)—consistently across experiments. Mostly, the mined model actually fits significantly worse.

We conclude that individual behavior is indeed best described by the memory-1

[1] Subjects randomize on actions every period, where randomization probabilities depend on actions in the precious period (memory-1) and cooperation probabilities after unilateral defection are independent of who defected.

behavior strategies that we consistently observe on average across experiments. Further, there is no evidence for heterogeneity aside from the usual 5–10% of noise players found in most experimental studies (i.e. subjects randomizing uniformly). In many ways, this represents good news for future research, as it shows that behavior is simple and robust, it confirms a central prediction, and implicitly the underlying assumptions, of recent theoretical analyses, and it poses a puzzling question for future theoretical and behavioral research: Why do subjects play the particular behavior strategies we consistently observe across this vast set of experiments? There is no obvious answer to this. In turn, our findings put a rather tight limitation on psychological interpretations of behavior. Individual behavior in various states of the repeated game is random, implying that causality between actions in one round and reactions in subsequent rounds is stochastic. In the most psychologically interesting states, succeeding unilateral defection, cooperation probabilities are around 30% and symmetric between cooperator and defector, rendering any causal statement linking action and reaction very weak.

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 $1 + g$ ($g > 0$). This constituent game is repeated infinitely often. Assuming players are risk neutral and discount future payoffs exponentially (using factor $\delta < 1$), this game is strategically equivalent to an indefinitely repeated one that is terminated with probability $1 - \delta$ after each round. We will refer to these games jointly as repeated prisoner's dilemma (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 the probability of cooperation in the next round. The strategy has *memory-1* if it prescribes the same cooperation probability for any two histories that do not differ in the actions chosen in their respective last rounds. It has memory-2 if the same holds for the respective last two rounds. We distinguish between the round-1 action and the continuation strategy, which prescribes behavior from round 2 on, and shall focus exclusively on analyzing continuation strategies. Thus, a memory-1 strategy may be denoted as $\sigma = (\sigma_{cc}, \sigma_{cd}, \sigma_{dc}, \sigma_{dd})$ corresponding to the four non-empty memory-1 histories $\{cc, cd, dc, dd\}$. We will refer to these as *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 strategy is a *pure strategy* if it prescribes degenerate cooperation probabilities after all histories ($\sigma \in \{0, 1\}$), and it is a *behavior strategy* otherwise. A *mixed strategy* randomizes on the set of pure strategies prior

to the start of the repeated game, in contrast to the behavior strategy that randomizes during the repeated game.

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, 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 1. A brief review is in Appendix B, but for a detailed discussion, see Dal Bó and Fréchette (2018). Thanks 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 purpose. 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 the hypotheses we intend to test. Finally, our analysis of continuation strategies complements the analysis by Dal Bó and Fréchette (2018), who focus largely on aggregate and first round cooperation rates as a function of treatment.

Table 1 provides an overview of behavior across experiments. It reports the average cooperation rates across experiments in each of the memory-1 states and tests for significance of difference. Besides clarifying average behavior, this allows us to test whether the surprising observation of Breitmoser (2015) that average strategies have a "semi-grim" pattern was specific to the four experiments he analyzed or can be considered a general phenomenon. A behavior strategy is called *semi-grim* if $\sigma_{cc} > \sigma_{cd} \approx \sigma_{dc} > \sigma_{dd}$, and indeed, this applies across most experiments, both for inexperienced subjects (the first half of sessions) and experienced subjects (second half of sessions). The differences between inexperienced and experienced subjects are clearly minor overall, the aggregate cooperation probabilities shift by at most five percentage points, but since it is customary to distinguish experienced and inexperienced behavior, we will consistently do so throughout this paper.

Our main hypothesis is that these semi-grim strategies reflect behavior not just on average, but also individually, as opposed to reflecting the sum of a variety of deterministic choice rules. The results of a very simple test of this hypothesis are reported in the last four columns of Table 1. 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 suffices to trigger significance in two-sided Fisher tests if subjects play a pure strategy. The results are fairly revealing: In state (c,d), i.e. after unilateral defection of the opponent, all standard pure strategies (except "always cooperate") 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

Table 1: Few subjects play pure strategies and assuming pure strategies yields a striking bias even in large mixture models

Experiment	Actual cooperation rates							Best-fitting rates assuming pure str.				Number of subjects not randomizing 50-50			
	$\hat{\sigma}_{cc}$		$\hat{\sigma}_{cd}$		$\hat{\sigma}_{dc}$		$\hat{\sigma}_{dd}$	$\tilde{\sigma}_{cc}$	$\tilde{\sigma}_{cd}$	$\tilde{\sigma}_{dc}$	$\tilde{\sigma}_{dd}$	(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	0.818^{-}	0.449	0.419	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	0.891	0.18^{-}	0.184	0.053	13/17	1/5	3/3	124/135
Bruttel and Kamecke (2012)	0.91	$\gg$	0.286	$\approx$	0.228	$\gg$	0.08	0.882	0.158^{--}	0.18	0.121	12/18	6/23	8/21	32/36
Dal Bó (2005)	0.922	$\gg$	0.212	$<$	0.342	$\gg$	0.089	0.896	0.16^{-}	0.346	0.108	13/13	0/3	2/2	42/54
Dal Bó and Fréchette (2011)	0.951	$\gg$	0.334	$\approx$	0.331	$\gg$	0.063	0.884^{-}	0.196^{--}	0.318	0.108	94/106	28/117	51/128	218/253
Dal Bó and Fréchette (2015)	0.94	$\gg$	0.297	$\approx$	0.335	$\gg$	0.057	0.891	0.172^{--}	0.317	0.09	216/243	37/137	62/147	404/474
Dreber et al. (2008)	0.904	$\gg$	0.217	$\approx$	0.213	$\gg$	0.036	0.915	0.081^{--}	0.199	0.084	15/25	3/19	12/18	45/48
Duffy and Ochs (2009)	0.904	$\gg$	0.301	$\approx$	0.33	$\gg$	0.111	0.863	0.239^{-}	0.342	0.15	43/57	4/25	10/24	61/82
Fréchette and Yuksel (2017)	0.943	$\gg$	0.141	$\approx$	0.266	$\approx$	0.091	0.918	0.099	0.274	0.085	21/28	0/0	2/2	5/8
Fudenberg et al. (2012)	0.982	$\gg$	0.4	$\approx$	0.427	$\gg$	0.066	0.946	0.295^{--}	0.404	0.06	38/43	1/6	5/11	20/25
Kagel and Schley (2013)	0.935	$\gg$	0.263	$\approx$	0.295	$\gg$	0.051	0.906	0.149^{--}	0.294	0.086	71/81	20/71	32/60	98/111
Sherstyuk et al. (2013)	0.945	$\gg$	0.328	$\approx$	0.371	$\gg$	0.117	0.868^{-}	0.201^{--}	0.377	0.125	37/44	10/36	12/34	41/52
Pooled	0.938	$\gg$	0.304	$\approx$	0.322	$\gg$	0.065	0.888	0.189^{--}	0.311	0.1	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	0.901^{-}	0.367	0.544	0.325	33/37	0/12	1/12	5/9
Blonski et al. (2011)	0.923	$\gg$	0.287	$\approx$	0.231	$\gg$	0.02	0.92	0.189^{-}	0.221	0.04	26/32	10/25	11/16	172/178
Bruttel and Kamecke (2012)	0.947	$\gg$	0.221	$\approx$	0.297	$\gg$	0.041	0.942	0.128^{-}	0.315	0.067	13/15	8/17	9/12	31/35
Dal Bó (2005)	0.92	$\gg$	0.242	$<$	0.388	$\gg$	0.064	0.914	0.193	0.386	0.1	18/27	0/3	0/1	50/65
Dal Bó and Fréchette (2011)	0.979	$\gg$	0.376	$\approx$	0.362	$\gg$	0.041	0.957	0.235^{--}	0.364	0.073	132/137	34/89	62/100	196/215
Dal Bó and Fréchette (2015)	0.976	$\gg$	0.315	$<$	0.402	$\gg$	0.035	0.947	0.187^{--}	0.408	0.061	340/365	52/162	77/146	448/497
Dreber et al. (2008)	0.917	$\gg$	0.128	$\ll$	0.39	$\gg$	0.009	0.936	0.087	0.395	0.044	14/18	6/11	6/12	41/43
Duffy and Ochs (2009)	0.977	$\gg$	0.367	$\approx$	0.391	$\gg$	0.082	0.923^{-}	0.205^{--}	0.37	0.085	80/87	5/35	16/43	60/68
Fréchette and Yuksel (2017)	0.97	$\gg$	0.233	$\approx$	0.398	$\gg$	0.069	0.93	0.072^{--}	0.391	0.074	33/37	1/6	2/10	20/25
Fudenberg et al. (2012)	0.971	$\gg$	0.487	$\approx$	0.412	$\gg$	0.083	0.942	0.43^{-}	0.378	0.091	41/44	2/8	4/10	14/17
Kagel and Schley (2013)	0.966	$\gg$	0.262	$\approx$	0.332	$\gg$	0.025	0.947	0.175^{-}	0.351	0.05	87/90	16/56	30/46	91/97
Sherstyuk et al. (2013)	0.973	$\gg$	0.482	$\approx$	0.437	$\gg$	0.078	0.919^{-}	0.375^{--}	0.4	0.122	44/48	7/24	17/23	23/29
Pooled	0.971	$\gg$	0.327	$<$	0.376	$\gg$	0.039	0.941	0.209^{--}	0.376	0.064	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). 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 "$\gg,\ll$". The "best-fitting rates assuming pure strategies" provide the cooperation probabilities explained assuming mixtures of pure strategies as usually analyzed in the literature (the five 1-memory and five 2-memory strategies usually considered). 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.

tight bound on the number of subjects potentially playing pure strategies.

The middle set of columns serves to illustrate the basic deficiency of deterministic choice rules. Assume that subjects use deterministic choice rules. Given that the semi-grim pattern results on average, there have to be subjects that systematically cooperate after unilateral defection of opponents. They do not have to cooperate always, but fairly often to make up for the general tendency toward defection in state (c,d) predicted by the standard strategies. For example, Result 6 of Dal Bó and Fréchette (2018) states that "always defect" (AD), Grim, and tit-for-tat (TFT) are the "three strategies [that] account for most of the data"—which seems to directly contradict the observation that $\sigma_{cd} \approx \sigma_{dc}$ unless around a third of the subjects systematically cooperate in state (c,d). Now allow for arbitrary mixtures of the ten pure strategies that had been found to be significant in previous analyses,[2] and let the mixture weights be adapted independently to each treatment of each experiment to optimally capture behavior. The middle set of columns compares the cooperation probabilities predicted by these best fitting mixtures (the econometric details follow in Section 3) with the actual cooperation probabilities by experiment. The quintessence is the strong bias in state (c,d), where predicted cooperation rates are substantially below the actual cooperation rates across experiments. Subjects cooperate much more often after unilateral defection of opponents than is compatible with the notion of pure strategies. The strategies predicting at least occasional cooperation after (c,d), such as always-cooperate and tit-for-2-tats, simply do not fit behavior of sufficiently many subjects to capture aggregate behavior. Based on this, we will obtain the range of highly significant results suggesting that subjects play behavior strategies reported below.

Related behavioral literature We keep the literature review short and focused due to the availability of the excellent recent survey by Dal Bó and Fréchette (2018). The modern experimental research on the repeated prisoner's dilemma started with Dal Bó (2005), who criticized earlier experiments for the implemented experimental designs such as letting 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 continuation strategies following round-1 that we also focus on. The general theme in the reported results is that initial cooperation rates depend on the strategic environment. More specifically, they show that subgame perfection of grim is necessary but not sufficient for cooperation to emerge, and that subsequent cooperation of

[2]These are five memory-1 strategies and five memory-2 strategies: tit-for-tat (TFT), grim, win-stay-lose-shift (WSLS), always defect (AD), always cooperate (AC), Grim2, tit-for-2-tats (TF2T), 2-tits-for-tat (2TFT), win-stay-lose-shift-2, and "T2" (i.e. punish defection for two periods, otherwise cooperate). The standard definitions are provided in Table 5 in the online appendix.

subjects depend on their opponent's actions, primarily on those in the previous round. Many of the second-wave analyses classify the strategies of individual subjects into varying pre-selected sets of pure strategies. These analyses typically conclude that the majority of subjects plays either AD, TFT, or Grim (Dal Bó and Fréchette, 2018, Result 6), with TFT being attributed population weights around 30% if the candidate set includes only pure strategies.

Breitmoser (2015) points out the misalignment of this claim with cooperation rates state-by-state. In a data set comprising four experiments he analyzes whether semi-grim behavior strategies better capture behavior than mixtures of pure memory-1 strategies, assuming each subject consistently plays a single memory-1 strategy throughout the experiment.[3] The results are suggestive, but obviously inconclusive, as the data set might be fortunately selected, as behavior might be more complex than memory-1 admits, and as subjects might switch strategies as the session progresses. In this paper, we report the results of an analysis relaxing all three concerns to arrive at arguably conclusive results. The data concern was addressed above. The case for memory-2 strategies had been made by Fudenberg et al. (2012), who show that if we assume subjects play pure strategies, then there must be subjects with memory-2, based on evidence for 2TFT and "forgiving" Grim2 strategies. Similar ideas are expressed in Aoyagi and Frechette (2009) and Bruttel and Kamecke (2012). In contrast to these studies, we relax the assumption of pure strategies in the first place, noting that behavior strategies themselves generate behavior resembling memory 2 or 3. Recently, Fréchette and Yuksel (2017) evaluate a pre-selected set of 20 memory-1 and memory-2/3 strategies and observe no evidence for memory-2/3 (in their random termination treatment). This observation does not contradict the earlier statements, as it may result due to the large set of candidate strategies considered in the first place, which amplifies standard errors, in particular if a single treatment is considered. Our analysis resolves such concerns using the data-mining techniques described in the next section.

3 Methodology and first results

Econometric approach 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 between supergames. A subject using a behavior strategy systematically deviates from pure strategies even within supergames. These definitions provide a basis for identification, given the set of pure strategies considered in the behavioral literature, but identification is made difficult by the standard assumption that choice is stochastic. Specifically, a single deviation from a given pure strategy, over say 20 observations, is intuitively not considered sufficient evidence against pu-

[3]The only other study investigating a behavior strategy seems to be Fudenberg et al. (2012), who include the strategy "generous TFT" which randomizes (only) after opponent's defection.

rity 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 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.[4] 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. The identifying assumption is that state transitions follow a Markov process, which substantially generalizes models assuming these transition probabilities to be degenerate (finite mixture) or uniform (random switching). 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 following Harless and Camerer (1994). All details are reported in Appendix A.

Data mining pure and mixed strategies Dal Bó and Fréchette (2018) argue that the five pure memory-1 strategies, namely AC, AD, TFT, Grim, and D-TFT, capture the behavior of most subjects across conditions. We focus on these strategies initially. Yet, as pure strategies have already shown to underestimate cooperation rates in state (c,d) in Table 1, we widen the scope of pure strategies by introducing generalized versions that allow for some randomization within supergames. Using the notation introduced above, generalized AC and AD are defined as behavior strategies $(\theta^{AC},\theta^{AC},\theta^{AC},\theta^{AC})$ and $(\theta^{AD},\theta^{AD},\theta^{AD},\theta^{AD})$ respectively, where θ^{AC} and θ^{AD} may be different. Generalized TFT is defined as $(1,0,\theta^{TFT},0)$, generalized Grim as $(1,\theta^{G},\theta^{G},\theta^{G})$, and generalized WSLS as $(1,0,0,\theta^{WSLS})$, with all $\theta^{*}\in[0,1]$. Given these definitions, we mine the data as follows. We illustrate all steps by referring to Table 2, which provides the results for memory-1 strategies.

First, we evaluate which mixture of pure or generalized pure strategies best captures behavior, independently for each treatment. That is, we determine for each treatment, which combination of *pure* strategies fits best, which combination of *generalized*

[4]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).

pure strategies fits best, and which of the combinations fits best overall. For simplicity, we assume the best combination always contains at least TFT, AD, and Grim, as those were reported to capture players' behavior in previous studies. This chooses the best out of the eight most promising memory-1 mixtures, for each of the 32 treatments independently. In total, we thus evaluate 8^{32} models per level of experience. Finally, we do all of this separately for three "switching models" designed to capture changes between supergames: "No Switching" (pure strategy), "Random Switching" (mixed strategy), and "Markov Switching" (where strategy choice between supergames follows a Markov process).

The results for each of the three switching models are reported in the first three columns of Table 2. For sake of readability, we aggregate ICL-BICs by experiment.[5] This initial analysis suggests that inexperienced subjects switch strategies randomly between supergames, while experienced subjects then play constant strategies. As a side note, we obtain strong evidence that subjects play the generalized pure strategies rather than the actual pure ones (Table 11 in the appendix provides details). The differences in model fit are large, amounting to more than 1000 points on the log-likelihood scale, suggesting that randomization within supergames is indeed a behavioral facet.

Second, given the best strategy combination at treatment level, we pick the best switching model by experience level, column 4 ("Best Switching") of Table 2, and evaluate this model against the simple semi-grim model reported in the fifth column. Note that we implicitly pick the best-fitting model from 3×8^{32} models, which is chosen after estimating 276 parameters per treatment, and evaluate it against the three-parametric behavior strategy semi-grim $(\theta_1^{SG}, \theta_2^{SG}, \theta_2^{SG}, \theta_3^{SG})$. In line with the data-mining ideal, we do not account for the degrees of freedom used in the model selection process, but solely account for the 3–10 parameters of the best-fitting model that is finally used: Still, the simple behavior strategy fits (weakly) better than the mined mixture of pure strategies. Below, we will verify if this observation holds similarly after accounting for memory-2 behavior.

Third, we make a first robustness check to clarify if perhaps not all subjects stick with memory-1 semi-grim strategies. We evaluate if some subjects simply randomize uniformly after all histories of play in the sense of level-0. The results are reported in the seventh column ("SG + Noise"). Allowing for noise players turns out to improve goodness-of-fit significantly for both experience levels—in line with many results in the literature (Nagel, 1995; Stahl and Wilson, 1995; Costa-Gomes et al., 2001), around 5-10 percent of the subjects are noise players. We conduct more such checks challenging memory-1 semi-grim below.

Fourth, we evaluate the full-blown model testing, independently treatment by treatment, which mixture and switching model fits behavior best. Thus, we choose the best-fitting model from 24 models for each treatment, amounting to the enormous

[5] Treatmentwise ICL-BICs are provided in the appendix, after Table 11. Each entry in the aggregated table represents the sum of ICL-BICs of the best out of eight models for each respective treatment.

Table 2: 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					
	No Switching		Random Switching		Markov Switching		Best Switching		Semi-Grim		SG + Noise		Best Switching By Treatment			
Specification																
# Models evaluated	8^{32}		8^{32}		8^{32}		3×8^{32}		1		1		$24^{32} \approx 10^{44}$			
# Pars estimated (by treatment)	48		48		180		276		3		4		276			
# Parameters accounted for	3–10		3–10		12-35		3–10		3		4		3–10			
First halves per session																
Aoyagi and Frechette (2009)	645.31	$\approx$	646.53	$\approx$	649.53		646.53	$\approx$	694.72	$>$	646.73	$\approx$	645.31			
Blonski et al. (2011)	677.42	$\gg$	619.43	$\ll$	821.29		619.43	$\gg$	549.45	$\approx$	558.66	$<$	614.47			
Bruttel and Kamecke (2012)	585.42	$\approx$	570.56	$\approx$	570.87		570.56	$\approx$	567.86	$\approx$	559.05	$\approx$	570.56			
Dal Bó (2005)	394.44	$\gg$	365.35	$\ll$	393		365.35	$\approx$	358.51	$\approx$	361.22	$\approx$	365.35			
Dal Bó and Fréchette (2011)	3536.73	$\approx$	3576.34	$\approx$	3524.77		3576.34	$\approx$	3533.99	$>$	3416.54	$\approx$	3471.98			
Dal Bó and Fréchette (2015)	5250.62	$\gg$	5006.33	$\approx$	5049.55		5006.33	$\approx$	4991.74	$\approx$	4935.93	$\approx$	4965.57			
Dreber et al. (2008)	459.11	$\approx$	463.01	$\approx$	477.02		463.01	$>$	437.17	$\approx$	435.65	$\approx$	461.11			
Duffy and Ochs (2009)	1047.59	$\approx$	1053.04	$\approx$	1049.79		1053.04	$\approx$	1090.22	$>$	1017.89	$\approx$	1047.59			
Fréchette and Yuksel (2017)	181.98	$\gg$	161.75	$<$	175.59		161.75	$\approx$	161.45	$\approx$	164.51	$\approx$	161.75			
Fudenberg et al. (2012)	319.45	$\approx$	308.6	$\approx$	320.55		308.6	$\approx$	291.43	$\approx$	288.89	$\approx$	308.6			
Kagel and Schley (2013)	1761.98	$\approx$	1780.97	$>$	1694.94		1780.97	$\approx$	1782.82	$\approx$	1726.06	$\approx$	1694.94			
Sherstyuk et al. (2013)	865.67	$\approx$	907.14	$>$	858.65		907.14	$\approx$	912.8	$\approx$	868.89	$\approx$	858.65			
Pooled	**15951.95**	$\gg$	**15675.49**	$\ll$	**16214.1**		**15675.49**	$>$	**15481.59**	$\gg$	**15125.92**	$\ll$	**15535.6**			
Second halves per session																
Aoyagi and Frechette (2009)	363.58	$\approx$	368.23	$\approx$	368.89		363.58	$\approx$	389.24	$>$	353.94	$\approx$	363.58			
Blonski et al. (2011)	946.2	$>$	912.16	$\ll$	1107.23		946.2	$>$	867.87	$\approx$	868.31	$\approx$	908.25			
Bruttel and Kamecke (2012)	342.17	$\approx$	358.12	$\approx$	347.08		342.17	$\approx$	347.4	$\approx$	342.6	$\approx$	342.17			
Dal Bó (2005)	461.23	$\approx$	445.57	$<$	469.98		461.23	$>$	424.44	$\approx$	429.06	$\approx$	442.59			
Dal Bó and Fréchette (2011)	2737.11	$<$	2865.45	$>$	2721.88		2737.11	$\approx$	2817.31	$\approx$	2694.4	$\approx$	2700.09			
Dal Bó and Fréchette (2015)	5153.82	$\approx$	5067.01	$\approx$	5106.57		5153.82	$\approx$	5043.81	$>$	4902.82	$\approx$	4998.09			
Dreber et al. (2008)	287.49	$\approx$	281.99	$\approx$	299.36		287.49	$\approx$	264.94	$\approx$	268.86	$\approx$	283.38			
Duffy and Ochs (2009)	1381.01	$\approx$	1416.71	$\approx$	1392.49		1381.01	$\approx$	1403.03	$>$	1308.8	$\approx$	1381.01			
Fréchette and Yuksel (2017)	309.63	$\approx$	304.7	$\approx$	308.78		309.63	$\approx$	313.5	$\approx$	278.74	$<$	304.7			
Fudenberg et al. (2012)	373.44	$\approx$	395.32	$\approx$	376.62		373.44	$\approx$	380.75	$\approx$	358.86	$\approx$	373.44			
Kagel and Schley (2013)	1170.12	$\approx$	1224.37	$>$	1143.67		1170.12	$\approx$	1211.37	$\approx$	1153.04	$\approx$	1143.67			
Sherstyuk et al. (2013)	527.09	$\approx$	590.16	$\approx$	567.63		527.09	$\approx$	586.72	$\approx$	564.5	$\approx$	527.09			
Pooled	**14269.01**	$\approx$	**14448.15**	$<$	**14877.81**		**14269.01**	$\approx$	**14159.8**	$\gg$	**13669.82**	$\ll$	**14108.4**			

Note: 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.

selection of the best out of 24^{32} models across all experiments. Note that such analysis without imposing consistency requirements across treatments does not yield economically useful estimates, but it provides an upper bound on the economic content of pure and generalized pure strategies. The results are reported in the seventh column ("Best Switching By Treatment"). In total, this exhaustively mined model still does not fit better than simple semi-grim strategy. It fits slightly worse than semi-grim for inexperienced subjects and slightly better for experienced ones, but all differences are insignificant. Yet, it fits significantly worse than semi-grim after controlling for the possibility that some subjects are pure noise players.

Result 1 (Memory-1)**.** *Assuming subjects have memory-1, the upper bound of behavior that can be captured after mining for pure or generalized pure strategies is indistinguishable from behavior captured by semi-grim, and significantly lower than by semi-grim accounting for noise players.*

4 Relaxing memory-1 and homogeneity

Data mining the memory-2 specification Next we extend the set of pure strategies to capture possible interdependence of actions with choices in $t-2$, again to evaluate the best fitting specification against memory-1 semi-grim. 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, e.g. , strategies that punish only after the second deviation or that punish for two rounds instead of one, respectively. This approach is applicable in particular to extend pure strategies, by providing a specific list of memory-2 generalizations. The other approach is parametric and is suitable in particular to extend generalized pure and behavior strategies from memory-1 to memory-2. It allows the cooperation probabilities in one round to depend on the behavior of one or both players in $t-2$. We allow for three different formulations here: 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*).

First, we mine for mixtures of pure strategies, based on the list of 10 strategies[6] of Fudenberg et al. (2012). We keep this description short, as the pure strategies fit similarly poorly as in the memory-1 case. For each treatment, we determine the most adequate combination of strategies from a list of five combinations, thus providing a selection of the best of 5^{32} models overall. The resulting model still fits highly significantly worse than the selection of generalized pure strategies with memory-1 defined above, see the two right-most columns of Table 3.

[6]These strategies are TFT, Grim, AD, Grim2, TF2T, T2, 2TFT, 2PTFT as defined in Fudenberg et al. (2012) and also in Table 5 in the Online Appendix.

Table 3: What is the memory length? Comparison of 1- and 2-memory Semi-Grim, pure and generalized strategies (ICL-BIC of the models, less is better and relation signs point toward better models)

	Memory-2 Generalizations of Semi-Grim						Semi-Grim		Best Mixtures of Generalized Pure Strategies						Best Pure
	M2"General"		M2"Grim"		M2"TFT"				M2"TFT"		M2"Grim"		M1		M1 & M2
Specification															
# Models evaluated	1		1		1		1		4^{32}		4^{32}		4^{32}		5^{32}
# Pars estimated (by treatment)	12		6		6		3		48		48		32		32
# Parameters accounted for	12		6		6		3		9–15		9–15		6–10		3–8
First halves per session															
Aoyagi and Frechette (2009)	692.5	$\approx$	690.85	$\approx$	686.2	$\approx$	694.72	$>$	649.23	$\approx$	646.7	$\approx$	645.31	$\ll$	791.38
Blonski et al. (2011)	714	$\gg$	601.67	$\approx$	601.95	$\gg$	549.45	$\ll$	760.49	$\approx$	767.03	$\gg$	713.8	$\approx$	703.1
Bruttel and Kamecke (2012)	572.14	$\approx$	566.75	$\approx$	567.58	$\approx$	567.86	$\approx$	577.54	$\approx$	581.45	$\approx$	585.42	$\approx$	588.55
Dal Bó (2005)	385.61	$>$	367.94	$\approx$	366.48	$\approx$	358.51	$\ll$	405.84	$\approx$	402.78	$\approx$	407.86	$\approx$	389.08
Dal Bó and Fréchette (2011)	3596.64	$\approx$	3542.28	$\approx$	3538.64	$\approx$	3533.99	$\approx$	3511.26	$\approx$	3527.68	$\approx$	3536.73	$\ll$	3835.75
Dal Bó and Fréchette (2015)	5017.27	$\approx$	4974.8	$\approx$	4988.94	$\approx$	4991.74	$\ll$	5269.9	$\approx$	5320.57	$\approx$	5259.64	$\ll$	5538.37
Dreber et al. (2008)	464.84	$>$	444.11	$\approx$	444.71	$\approx$	437.17	$\approx$	477.33	$\approx$	483.11	$\approx$	478.09	$\approx$	462.71
Duffy and Ochs (2009)	1060.26	$\approx$	1063.66	$\approx$	1074.9	$\approx$	1090.22	$\approx$	1036.79	$\approx$	1044.14	$\approx$	1047.59	$\approx$	1102.64
Fréchette and Yuksel (2017)	174.64	$\approx$	167.06	$\approx$	164.75	$\approx$	161.45	$\ll$	182.81	$\approx$	188.15	$\approx$	188.5	$\approx$	181.98
Fudenberg et al. (2012)	301.76	$\approx$	293.52	$\approx$	294.4	$\approx$	291.43	$<$	319.76	$\approx$	322.36	$\approx$	319.45	$<$	366.78
Kagel and Schley (2013)	1746.26	$\approx$	1749.95	$\approx$	1753.68	$\approx$	1782.82	$>$	1634.47	$\approx$	1665.36	$<$	1761.98	$\approx$	1805.95
Sherstyuk et al. (2013)	917.07	$\approx$	907.95	$\approx$	913.52	$\approx$	912.8	$\approx$	872.66	$\approx$	873.21	$\approx$	865.67	$<$	941.91
Pooled	**16080.69**	$\gg$	**15589.39**	$\approx$	**15614.59**	$>$	**15481.59**	$\ll$	**16083.49**	$\approx$	**16207.37**	$\approx$	**16077.95**	$\ll$	**16858.86**
Second halves per session															
Aoyagi and Frechette (2009)	396.32	$\approx$	391.42	$>$	387.48	$\approx$	389.24	$\approx$	368.09	$\approx$	365.6	$\approx$	363.58	$\ll$	484.41
Blonski et al. (2011)	1012.48	$\gg$	919.29	$\approx$	922.48	$\gg$	867.87	$<$	1005.42	$\approx$	1020.36	$>$	992.44	$\approx$	1055.95
Bruttel and Kamecke (2012)	333.51	$\approx$	337.12	$\approx$	329.73	$\approx$	347.4	$\approx$	324.09	$\approx$	336.35	$\approx$	344.88	$\approx$	316.38
Dal Bó (2005)	449.03	$\approx$	434.38	$\approx$	433.82	$\approx$	424.44	$<$	451.34	$<$	471.65	$\approx$	475.11	$\approx$	463.54
Dal Bó and Fréchette (2011)	2854.52	$\approx$	2801.46	$\approx$	2800.71	$\approx$	2817.31	$\approx$	2664.52	$\approx$	2690.03	$\approx$	2737.11	$<$	2885.43
Dal Bó and Fréchette (2015)	5006.3	$\approx$	5013.49	$\approx$	5012.99	$\approx$	5043.81	$\approx$	5100.98	$<$	5202.24	$\approx$	5164.78	$\ll$	5577.55
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	$\approx$	1348.96	$\approx$	1367.92	$\approx$	1381.01	$\ll$	1617.77
Fréchette and Yuksel (2017)	308.21	$\approx$	304.2	$\approx$	306.93	$\approx$	313.5	$\approx$	311.04	$\approx$	311.99	$\approx$	309.63	$\ll$	356.11
Fudenberg et al. (2012)	384.37	$\approx$	382.32	$\approx$	378.59	$\approx$	380.75	$\approx$	364.71	$\approx$	359.37	$\approx$	373.44	$<$	447.19
Kagel and Schley (2013)	1204.38	$\approx$	1202.61	$\approx$	1197.19	$\approx$	1211.37	$>$	1052.18	$\approx$	1066.23	$\ll$	1170.12	$\approx$	1169.31
Sherstyuk et al. (2013)	598.79	$\approx$	590.65	$\approx$	591.38	$\approx$	586.72	$>$	510.29	$\approx$	523.29	$\approx$	527.09	$\approx$	583.8
Pooled	**14633.97**	$\gg$	**14222.35**	$\approx$	**14223.53**	$\approx$	**14159.8**	$\approx$	**14152.67**	$\ll$	**14384.17**	$\approx$	**14387.48**	$\ll$	**15402.36**

Note: The main body contains ICL-BICs aggregated at paper level. Relation signs and p-values are exactly as above, see Table 2. "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 5 in the Online Appendix.

Second, we evaluate extensions of the generalized pure strategies to memory-2, using both the TFT-scheme and the Grim-scheme as defined above. That is, the strategy parameters $(\theta^{AC}, \theta^{AD}, \theta^{TFT}, \theta^{G}, \theta^{WSLS})$ are allowed to be functions of the opponent's action in $t-2$ (TFT-scheme) or of whether both players cooperated in $t-2$ (Grim-scheme). For each treatment, we determine the best of four combinations exactly as in the memory-1 case, thus determining the best of 4^{32} specifications overall. The results are reported in three columns headlined "Best Mixtures of Generalized Pure Strategies" in Table 3. In general, the TFT-scheme seems to fit better than the Grim-scheme, and overall, allowing for memory-2 helps to better capture behavior only if subjects are experienced. This suggests that behavior becomes more nuanced as subjects gain experience, but the resulting model does still not improve on the simple semi-grim strategy. Note that this obtains despite the enormous number of free parameters used in the estimation, 48 by treatment instead of 3—the pure and generalized pure strategies plainly have the wrong structure, for the reasons outlined already in section 2.

Third, we proceed similarly for evaluating extensions of semi-grim toward memory-2. That is, we compare the simple three-parametric memory-1 version with three generalizations to memory-2. As 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 again 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. Let us summarize this as follows.

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

As a corollary, Result 1 on memory-1 behavior appears to be a generally adequate statement. In order to further verify its adequacy, we analyze if behavior of residual subjects is better captured by non-trivial strategies than by the simple Noise strategy from above. That is, are subject pools in the repeated prisoner's dilemma significantly heterogeneous beyond the existence of noise players? Data-mining exactly as above, we evaluate a wide range of specifications to capture behavior of the residual players, in order to obtain an upper bound of model adequacy for classes of specifications, and evaluate the best possible combination of specifications across treatments against the simple alternative that these residual players are noise players.

First, we evaluate the adequacy of pure or generalized pure strategies to complement semi-grim. Specifically, we evaluate finite-mixture models combining semi-grim with each of the five pure memory-1 strategies (AD, AC, TFT, WSLS, and Grim) or each of the four generalized pure strategies (generalized AD and generalized AC are

Table 4: Heterogeneity beyond semi-grim: Is there are significant secondary subject type?

	Mixtures of SG with pure/generalized pure						SG + Noise		Mixing Semi-Grim with					
	Best Pure		Best Gen		Pure or Gen				SG		SG-M2"Grim"		SG-M2"TFT"	
Specification														
# Models evaluated	5^{32}		4^{32}		9^{32}		1		1		1		1	
# Pars estimated (by treatment)	20		20		40		4		7		10		10	
# Parameters accounted for	4		5		4–5		4		7		10		10	
First halves per session														
Aoyagi and Frechette (2009)	681.99	$\approx$	649.86	$\approx$	649.86	$\approx$	646.73	$\approx$	650.83	$\approx$	651.47	$\approx$	656.3	
Blonski et al. (2011)	576.49	$\approx$	600.48	$\gg$	568.43	$\approx$	558.66	$\ll$	642.5	$\ll$	711.05	$\approx$	717.67	
Bruttel and Kamecke (2012)	562.35	$\approx$	559.64	$\approx$	559.64	$\approx$	559.05	$\approx$	553.59	$\approx$	555.66	$\approx$	553.75	
Dal Bó (2005)	367.32	$\approx$	363.9	$\approx$	363.9	$\approx$	361.22	$<$	379.63	$\approx$	388.26	$\approx$	390.43	
Dal Bó and Fréchette (2011)	3413.18	$\approx$	3379.17	$\approx$	3368.28	$\approx$	3416.54	$\approx$	3335.4	$\approx$	3386.94	$\approx$	3375.88	
Dal Bó and Fréchette (2015)	4950.44	$\approx$	4920.88	$\approx$	4905.86	$\approx$	4935.93	$\approx$	4958.73	$\approx$	4990.23	$\approx$	5005.61	
Dreber et al. (2008)	432.72	$\approx$	439.1	$\approx$	432.72	$\approx$	435.65	$\approx$	438.35	$\approx$	443.27	$\approx$	441.52	
Duffy and Ochs (2009)	1074.22	$>$	1017.65	$\approx$	1017.65	$\approx$	1017.89	$\approx$	1016.91	$\approx$	1023.24	$\approx$	1016.46	
Fréchette and Yuksel (2017)	165.36	$\approx$	167.94	$\approx$	165.36	$\approx$	164.51	$<$	179.6	$\approx$	181.01	$\approx$	181.57	
Fudenberg et al. (2012)	292.47	$\approx$	291.68	$\approx$	291.68	$\approx$	288.89	$\approx$	287.31	$<$	308.85	$\approx$	308.99	
Kagel and Schley (2013)	1685.56	$\approx$	1690.55	$\approx$	1685.56	$\approx$	1726.06	$>$	1677.99	$\approx$	1655.61	$\approx$	1639.86	
Sherstyuk et al. (2013)	887.79	$\approx$	866.46	$\approx$	866.46	$\approx$	868.89	$\approx$	868.11	$\approx$	877.05	$\approx$	877.26	
Pooled	**15272.25**	$\approx$	**15166.16**	$\approx$	**15078.27**	$\approx$	**15125.92**	$\approx$	**15244.28**	$\ll$	**15537.4**	$\approx$	**15530.04**	
Second halves per session														
Aoyagi and Frechette (2009)	392.87	$>$	357.21	$\approx$	357.21	$>$	353.94	$\approx$	358.14	$\approx$	358.14	$\approx$	359.37	
Blonski et al. (2011)	861.53	$\approx$	876.85	$\gg$	855.06	$\approx$	868.31	$\approx$	906.87	$\ll$	978.33	$\approx$	982.05	
Bruttel and Kamecke (2012)	336.68	$>$	323.04	$\approx$	323.04	$\approx$	342.6	$\approx$	322.87	$\approx$	314.07	$\approx$	308.75	
Dal Bó (2005)	431.34	$\approx$	435.55	$\approx$	431.34	$\approx$	429.06	$\ll$	455.9	$\ll$	478.45	$\approx$	467.72	
Dal Bó and Fréchette (2011)	2597.41	$\approx$	2543.05	$\approx$	2537.58	$\ll$	2694.4	$>$	2545.23	$\approx$	2508.92	$\approx$	2556.33	
Dal Bó and Fréchette (2015)	4946.41	$\approx$	4884.99	$\approx$	4868.35	$\approx$	4902.82	$\approx$	4879.57	$\approx$	4882.52	$\approx$	4958.15	
Dreber et al. (2008)	262.93	$\approx$	258.73	$\approx$	256.77	$\approx$	268.86	$\approx$	268.1	$\approx$	252.76	$\approx$	254.09	
Duffy and Ochs (2009)	1379.73	$>$	1294.4	$\approx$	1294.4	$\approx$	1308.8	$\approx$	1284.59	$\approx$	1278.88	$\approx$	1290.71	
Fréchette and Yuksel (2017)	311.64	$\approx$	283.31	$\approx$	283.31	$\approx$	278.74	$\approx$	277.03	$\approx$	278.37	$\approx$	274.84	
Fudenberg et al. (2012)	366.79	$\approx$	364.37	$\approx$	364.37	$\approx$	358.86	$\approx$	348.86	$\approx$	351.64	$\approx$	352.94	
Kagel and Schley (2013)	1126.89	$\approx$	1154.97	$\approx$	1126.89	$\approx$	1153.04	$\approx$	1111.19	$\approx$	1086.76	$\approx$	1084.65	
Sherstyuk et al. (2013)	539.35	$\approx$	565.48	$\approx$	539.35	$\approx$	564.5	$\approx$	544.85	$\approx$	543.88	$\approx$	538.39	
Pooled	**13735.95**	$\approx$	**13560.79**	$\approx$	**13438.59**	$<$	**13669.82**	$\approx$	**13558.53**	$\approx$	**13677.46**	$\approx$	**13792.73**	

Note: The main body contains ICL-BICs aggregated at paper level. The first three columns pick the best fitting combinations of SG with pure, generalized, and generalized or pure memory-1 strategies. The last three columns combine memory-1 SG with a second memory-1 SG strategy and two different versions of a memory-2 SG strategy respectively. As above, the 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.

equivalent here). We then determine which of these models best captures behavior, treatment by treatment, to identify upper bounds of model adequacy for augmenting semi-grim with pure or generalized pure strategies. Table 4 presents the results. The first column "Best Pure" refers to the best treatment-wise combinations of semi-grim with pure strategies, "Best Gen" refers to combinations with generalized pure strategies, and "Pure or Gen" refers to the best combination with pure or generalized pure strategies. Doing so treatment by treatment, this yields a selection from 9^{32} models. Comparing aggregate ICL-BICs of the mined model to those containing noise players as complement to semi-grim indicates that there is little scope for improvement in this direction. There is no significant difference in model fit if subjects are inexperienced, and there is only one experiment for which the combination of semi-grim with a (generalized) pure strategy performs significantly better than SG+Noise. For this one experiment the best mixture is a combination of semi-grim with generalized AD/AC, which is very similar to noise: It represents unconditional randomization between cooperation and defection but allows average cooperation rates to deviate from 0.5 (at the expense of a degree of freedom).

Second, we evaluate if a residual component is better captured using a memory-1 or memory-2 semi-grim (SG) strategy instead of noise. As memory-2 implementations we evaluate the TFT- and Grim-schemes introduced above. The overall picture can be summarized succinctly: Overall, none of these models improves on the null that the residual players simply randomize fifty-fifty. The results are presented in the three right-most columns of Table 4. The aggregate ICL-BICs of the two models combining semi-grim with memory-2 semi-grim do not significantly differ from each other, both perform slightly worse than the simpler model combining memory-1 SG with another memory-1 SG ("SG+SG"), which in turn does not perform better than the SG+Noise model. Since all of these semi-grim models contain uniform randomization (noise) as special case, we conclude that the additional structure provided by these models does not improve the goodness-of-fit sufficiently to warrant the free parameters used in the process. Noise is an adequate description for the residual players.

Result 3. *There is significant evidence for subject heterogeneity, but no single model of behavior significantly improves on the hypothesis that the residual component besides semi-grim subjects consists simply of "noise players" randomizing fifty-fifty.*

5 Conclusion

Do we play pure, mixed, or behavior strategies in the repeated prisoner's dilemma? What is the memory length of these strategies and are subjects heterogeneous? After data mining for the best-fitting mixtures of pure strategies and switching models, which puts pure and mixed strategies at an enormous advantage, we still have to reject the hypothesis that subjects play mixtures of deterministic strategies in favor of the much simpler explanation that subjects actually play the strategies that we observe

consistently on average. Even in our large data set, we find no evidence for the use of memory-2 strategies nor for subject heterogeneity beyond the existence of noise players. That is, aggregating across 12 experiments and 145,000 decisions, subjects play memory-1 behavior strategies highly consistently. These strategies depend to some extent on the treatment parameters, see Table 1, and understanding this interdependence is thus established as the key to understanding cooperation in long-run relationships.

This suggests that subjects decide "just in time" whether to retaliate or be lenient. From the position of an observer, this is best described as a random Markov process (a memory-1 behavior strategy), as opposed to a complex set of deterministic rules. From the position of the decision maker herself, any decision may appear to be perfectly deterministic, but this discrepancy has been labeled the "illusion of conscious will" (Wegner, 2004) as any decision is preceded by unconscious cerebral activity feeding us thoughts (Libet, 1985). From a game-theoretic perspective, individual decisions are not sufficiently consistent to organize them by means of deterministic rules, but an illusion of conscious will would be reminiscent of purification in the sense of Harsanyi (1973).

Our results clarify which statistics to analyze—memory-1 cooperation rates—in experimental studies of long-run interactions and underscore the practical relevance of the current theoretical literature on imperfect private monitoring, which actually predicts that players would play simple behavior strategies. The strategies observed in our analysis are very specific instances of belief-free equilibria, however, played consistently across a wide range of experiments and treatments, which is not an obvious implication of the theory as it is currently developed. The hypothesis that equilibria are additionally robust to utility perturbations (Breitmoser, 2015) may be worth investigating, but it should be noted that the cooperation probabilities exhibit semi-grim patterns even if belief-free equilibria sustaining cooperation do not exist. Analogical reasoning in the sense of Samuelson (2001) may help explain these "anomalies", but more generally, the consistency of the semi-grim pattern in cooperation probabilities seems to represent a non-trivial challenge for future research—despite the already impressive body of theoretical work on repeated games.

References

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.

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., 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.

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.

Ely, J. C. and Välimäki, J. (2002). A robust folk theorem for the prisoner's dilemma. *Journal of Economic Theory*, 102(1):84–105.

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.

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 Levine, D. K. (1993). Self-confirming equilibrium. *Econometrica*, 61(3):523–545.

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.

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

Harsanyi, J. C. (1973). Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory*, 2(1):1–23.

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.

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

Kandori, M. (2002). Introduction to repeated games with private monitoring. *Journal of Economic Theory*, 102(1):1–15.

Kandori, M. (2011). Weakly belief-free equilibria in repeated games with private monitoring. *Econometrica*, 79(3):877–892.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and brain sciences*, 8(4):529–539.

Mailath, G. J. and Samuelson, L. (2006). *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review*, 85(5):1313–1326.

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.

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

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

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.

Stahl, D. O. and Wilson, P. W. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254.

Wegner, D. M. (2004). Précis of the illusion of conscious will. *Behavioral and Brain Sciences*, 27(5):649–659.

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

God does not play dice, but do we?

Appendix A: Markov-Switching Models

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

$$\ln \mathcal{L}(\theta, \rho | O) = \log \left(\prod_{s \in S} p(o_s | \theta, \rho) \right) = \sum_{s \in S} \ln \sum_{k \in K} \rho_k \, p_k(o_s | \theta_k), \tag{1}$$

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 [7].

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.[8] 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

[7] 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 $\omega_{s,t}(k)^{1-o_{s,t}}$.

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

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:

$$\ln \mathcal{L}(\rho,\sigma,\theta|O) = \sum_{s\in S} \ln \sum_{\kappa\in K^T} Pr(\kappa) \prod_{t\leq T} Pr(o_{s,t}|\theta,\kappa(t),t) \tag{2}$$

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:

$$\alpha_{sk}(t) = Pr(O_{s1} = o_{s1},...,O_{st} = o_{st}, K_t = k) \tag{3}$$

Recursively, we can then define:

$$\begin{aligned}
&1. \quad \alpha_{sk}(1) = \rho_k Pr(o_{s1}|\theta,k) \qquad (4)\\
&2. \quad \alpha_{sk'}(t+1) = \left[\sum_k \alpha_{sk}(t)\sigma_{kk'}\right] Pr(o_{st+1}|\theta,k')\\
&3. \quad Pr(o_s) = \sum_{k\in K} \alpha_{sk}(T)
\end{aligned}$$

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.

$$\beta_{sk}(t) = Pr(O_{st+1} = o_{st+1},...,O_T = o_T|K_t = k) \tag{5}$$

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

$$\begin{aligned} &1. \quad \beta_{sk}(T) = 1 \\ &2. \quad \beta_{sk}(t) = \sum_{k' \in K} \sigma_{kk'} Pr(o_{st+1}|\theta, k')\beta_{sk'}(t+1) \\ &3. \quad Pr(o_s) = \sum_{k \in K} \beta_{sk}(1)\rho_k Pr(o_{s1}|\theta, k) \end{aligned} \tag{6}$$

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:

$$\gamma_{sk}(t) = Pr(K_t = k|o_s) = \frac{Pr(o_s, K_t = k)}{Pr(o_s)} = \frac{Pr(o_s, K_t = k)}{\sum_{k' \in K} Pr(o_s, K_t = k)} = \frac{\alpha_{sk}(t)\beta_{sk}(t)}{\sum_{k' \in K} \alpha_{sk'}(t)\beta_{sk'}(t)}, \tag{7}$$

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

$$\begin{aligned} \zeta_{skk'}(t) = Pr(K_t = k, K_{t+1} = k'|o_s) &= \frac{Pr(K_t = k|o_s)Pr(o_{t+1,\ldots,T}, K_{t+1} = k'|K_t = k')}{Pr(o_{t+1,\ldots,T}|K_t = k)} \\ &= \frac{\gamma_{sk}(t)\sigma_{kk'}Pr(o_{s,t+1}|\theta, k')\beta_{sk'}(t+1)}{\beta_{sk}(t)} \end{aligned} \tag{8}$$

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

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

$$LL_{kt}(\theta'_k) = \sum_{s \in S} \gamma_{sk}(t) \ln Pr(o_{st}|\theta') \rightarrow \max_{\theta'_{kt}} ! \tag{9}$$

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}.[9] Moreover, we update ρ and σ using the posteriors from above and yield

$$\rho_k^{+1} = \frac{1}{n} \sum_{s \in S} \gamma_{sk}(1) \quad \text{and} \quad \sigma_{kk'}^{+1} = \frac{\sum_{s \in S} \sum_{t<T} \zeta_{skk'}(t)}{\sum_{s \in S} \sum_{t<T} \gamma_{sk}(t)} \tag{10}$$

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

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

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 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.

Appendix B: Information on the experiments re-analyzed

This section provides some background information on the experiments re-analyzed in this paper. Table 5 summarizes and defines the strategies considered by previous studies. Table 6 reviews focus and main results (in terms of identified strategies) of these studies. Table 7 reviews the numbers of subjects and observations, average parameters, and average cooperation rates for all experiments, and Table 8 provides the detailed overview by treatments.

Table 5: 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)	DF11, DF15, FRD12, 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 6: 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 7: 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 8: Table 7 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

Appendix C: Robustness checks for Section 3

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 9 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 focusing 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 11 compares the best mixtures of pure and generalized pure strategies as discussed in the main text.
- Table 13 is similar to Table 2 in the main text but focusing on the prototypical strategies in their pure form.
- Table 15 is similar to Table 2 in the main text but focusing on the prototypical strategies in their generalized form.
- Table 17 is equivalent to Table 2 in the main text.
- Table 19 examines a number of mixture models involving Semi-Grim strategies to clarify that the mixtures analyzed in Table 2 do neither misrepresent the general picture (these alternative models are not better in terms goodness-of-fit and hence need not be discussed in detail in the main text) nor present a non-robust part of the picture (these alternative models do not fit substantially worse than simple Semi-Grim either, implying that the general results on the relation of Semi-Grim to the prototypical strategies can be considered robust).

Table 9: 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	4		4		4		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)	790.2	$>$	767.22	$\approx$	797.89		790.2	$>$	767.22	$\approx$	797.89	
Blonski et al. (2011)	674.3	$\gg$	626.71	$\ll$	843		719.21	$\gg$	678.16	$\ll$	1230.45	
Bruttel and Kamecke (2012)	590.68	$\approx$	581	$\approx$	592.44		590.68	$\approx$	582.82	$\approx$	601.83	
Dal Bó (2005)	394.44	$\gg$	365.35	$\ll$	393		412.2	$\gg$	376.98	$\ll$	472.2	
Dal Bó and Fréchette (2011)	3847.44	$\approx$	3775.96	$\approx$	3802.38		3872.8	$\approx$	3785.92	$\ll$	3927.78	
Dal Bó and Fréchette (2015)	5522.08	$\gg$	5280.06	$\approx$	5318.33		5549.01	$\gg$	5332.22	$\ll$	5531.46	
Dreber et al. (2008)	459.11	$\approx$	462.48	$\approx$	481.64		464.92	$\approx$	466.28	$\ll$	532.59	
Duffy and Ochs (2009)	1128.03	$\approx$	1111.23	$\approx$	1123.74		1128.79	$\approx$	1111.23	$\approx$	1134.58	
Fréchette and Yuksel (2017)	181.98	$\gg$	161.75	$<$	176.5		186.7	$>$	163.08	$\ll$	199.13	
Fudenberg et al. (2012)	356.73	$>$	331.44	$<$	347.07		359.72	$>$	333.38	$\ll$	366.32	
Kagel and Schley (2013)	1813.04	$\approx$	1862.88	$\gg$	1781.74		1815.98	$\approx$	1865.9	$>$	1790.48	
Sherstyuk et al. (2013)	926.9	$\approx$	953.91	$>$	912.67		928.91	$\approx$	955.58	$>$	922.05	
Pooled	**16820.09**	$\gg$	**16417.97**	$\ll$	**17165.74**		**17001.48**	$\gg$	**16601.15**	$\ll$	**18601.02**	
Second halves per session												
Aoyagi and Frechette (2009)	478.46	$>$	451.68	$\approx$	480.65		480.97	$>$	451.68	$<$	491.85	
Blonski et al. (2011)	965.51	$\gg$	914.28	$\ll$	1140.5		1005.4	$\gg$	956.4	$\ll$	1503.15	
Bruttel and Kamecke (2012)	342.17	$\approx$	361.38	$\approx$	348.88		343.97	$\approx$	362.67	$\approx$	368.05	
Dal Bó (2005)	464.42	$\approx$	446.63	$\ll$	474.71		474.92	$>$	453.64	$\ll$	555.44	
Dal Bó and Fréchette (2011)	3052.84	$\approx$	3081.92	$\approx$	3092.35		3069.11	$\approx$	3096.21	$<$	3203.86	
Dal Bó and Fréchette (2015)	5571.71	$>$	5419.19	$\approx$	5478.11		5587.96	$>$	5434.22	$\ll$	5637.28	
Dreber et al. (2008)	287.58	$\approx$	285.34	$<$	303.79		295.44	$\approx$	293.01	$\ll$	370.54	
Duffy and Ochs (2009)	1628.02	$\approx$	1628.89	$\approx$	1609.73		1630.32	$\approx$	1628.89	$\approx$	1621.54	
Fréchette and Yuksel (2017)	356.12	$\approx$	332.38	$<$	344.23		362.94	$>$	333.17	$\ll$	370.49	
Fudenberg et al. (2012)	443.13	$\approx$	439.28	$\approx$	444.41		445.07	$\approx$	441.4	$\approx$	463.88	
Kagel and Schley (2013)	1213.26	$<$	1301.76	$\gg$	1203.05		1215.63	$<$	1303.97	$>$	1223.4	
Sherstyuk et al. (2013)	587.45	$<$	640.1	$>$	597.28		588.7	$<$	640.41	$>$	609.01	
Pooled	**15523.37**	$\approx$	**15443.11**	$\ll$	**16124.95**		**15682.79**	$\approx$	**15578.03**	$\ll$	**17512.74**	

Note: Relation signs are used as defined above (Table 7). "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 10: Table 9 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	4		4		4		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	790.2	$>$	767.22	$\approx$	797.89		790.2	$>$	767.22	$\approx$	797.89	
BOS11–9	45.36	$\approx$	43.78	$\ll$	55.54		46.95	$\approx$	47.98	$\ll$	82.02	
BOS11–14	61.42	$\approx$	59.77	$\ll$	72.93		64.41	$\approx$	62.85	$\ll$	99.82	
BOS11–15	29.87	$\approx$	29.93	$\ll$	43.24		32.88	$\approx$	33.11	$\ll$	70.23	
BOS11–16	106.92	$\approx$	103.63	$\ll$	118.03		108.38	$\approx$	104.63	$\ll$	139.03	
BOS11–17	48.66	$<$	53.62	$\approx$	57.44		54.66	$\approx$	54.85	$\ll$	82.08	
BOS11–26	162.85	$\gg$	140.64	$\ll$	153.41		171.26	$\gg$	142.95	$\ll$	182.93	
BOS11–27	65.77	$>$	52.93	$\ll$	65.67		67.27	$\approx$	67.14	$\ll$	89.62	
BOS11–30	30.38	$\approx$	30.39	$\ll$	43.76		33.4	$\approx$	33.5	$\ll$	70.72	
BOS11–31	88.41	$>$	79.66	$\ll$	103.59		89.91	$>$	81.06	$\ll$	113.56	
BK12–28	590.68	$\approx$	581	$\approx$	592.44		590.68	$\approx$	582.82	$\approx$	601.83	
D05–18	153.77	$\gg$	136.89	$\ll$	151.01		158.41	$\gg$	141.54	$\ll$	182.66	
D05–19	238.55	$\approx$	226.34	$\approx$	233.48		250.24	$>$	231.9	$\ll$	268.27	
DF11–6	379.95	$>$	357.02	$\approx$	347.58		382.33	$>$	357.08	$\approx$	359.31	
DF11–7	549.14	$>$	513.5	$\approx$	514.91		551.1	$>$	516.11	$\approx$	528.46	
DF11–8	581.17	$\approx$	568.86	$\approx$	552.52		583.54	$\approx$	570.76	$\approx$	562.3	
DF11–22	758.46	$\approx$	738.85	$\approx$	742.48		763.64	$\approx$	740.75	$\approx$	757.73	
DF11–23	715.77	$\approx$	735.67	$\approx$	706.54		717.59	$\approx$	735.67	$\approx$	718.8	
DF11–24	843.2	$\approx$	838.63	$\approx$	830.64		847.67	$\approx$	838.63	$\approx$	839.63	
DF15–4	170.82	$\approx$	158.62	$\approx$	164.65		175.88	$\approx$	162.97	$\ll$	198.4	
DF15–5	761.26	$\approx$	724.76	$\approx$	696.32		765.29	$\approx$	729.13	$\approx$	717.94	
DF15–20	1003.49	$>$	955.28	$\approx$	958.24		1005.86	$>$	967.96	$\approx$	973.01	
DF15–21	1334.08	$\approx$	1316.87	$\gg$	1274.95		1336.6	$\approx$	1333.38	$\gg$	1296.16	
DF15–33	1900.51	$\gg$	1796.11	$\approx$	1810.13		1904.58	$\gg$	1800.3	$<$	1831.38	
DF15–35	330.58	$>$	307.28	$\approx$	319.5		331.62	$>$	309.31	$\ll$	339.6	
DRFN08–10	197.04	$\approx$	195.4	$\approx$	208.99		198.7	$\approx$	197.07	$\ll$	229.14	
DRFN08–11	259.69	$\approx$	264.28	$\approx$	264.24		262.72	$\approx$	265.71	$\approx$	282.44	
DO09–32	1128.03	$\approx$	1111.23	$\approx$	1123.74		1128.79	$\approx$	1111.23	$\approx$	1134.58	
FY17–25	181.98	$\gg$	161.75	$<$	176.5		186.7	$>$	163.08	$\ll$	199.13	
FRD12–29	356.73	$>$	331.44	$<$	347.07		359.72	$>$	333.38	$\ll$	366.32	
KS13–12	1813.04	$\approx$	1862.88	$\gg$	1781.74		1815.98	$\approx$	1865.9	$>$	1790.48	
STS13–13	926.9	$\approx$	953.91	$>$	912.67		928.91	$\approx$	955.58	$>$	922.05	
Aoyagi and Frechette (2009)	790.2	$>$	767.22	$\approx$	797.89		790.2	$>$	767.22	$\approx$	797.89	
Blonski et al. (2011)	674.3	$\gg$	626.71	$\ll$	843		719.21	$\gg$	678.16	$\ll$	1230.45	
Bruttel and Kamecke (2012)	590.68	$\approx$	581	$\approx$	592.44		590.68	$\approx$	582.82	$\approx$	601.83	
Dal Bó (2005)	394.44	$\gg$	365.35	$\ll$	393		412.2	$\gg$	376.98	$\ll$	472.2	
Dal Bó and Fréchette (2011)	3847.44	$\approx$	3775.96	$\approx$	3802.38		3872.8	$\approx$	3785.92	$\ll$	3927.78	
Dal Bó and Fréchette (2015)	5522.08	$\gg$	5280.06	$\approx$	5318.33		5549.01	$\gg$	5332.22	$\ll$	5531.46	
Dreber et al. (2008)	459.11	$\approx$	462.48	$\approx$	481.64		464.92	$\approx$	466.28	$\ll$	532.59	
Duffy and Ochs (2009)	1128.03	$\approx$	1111.23	$\approx$	1123.74		1128.79	$\approx$	1111.23	$\approx$	1134.58	
Fréchette and Yuksel (2017)	181.98	$\gg$	161.75	$<$	176.5		186.7	$>$	163.08	$\ll$	199.13	
Fudenberg et al. (2012)	356.73	$>$	331.44	$<$	347.07		359.72	$>$	333.38	$\ll$	366.32	
Kagel and Schley (2013)	1813.04	$\approx$	1862.88	$\gg$	1781.74		1815.98	$\approx$	1865.9	$>$	1790.48	
Sherstyuk et al. (2013)	926.9	$\approx$	953.91	$>$	912.67		928.91	$\approx$	955.58	$>$	922.05	
Pooled	**16820.09**	$\gg$	**16417.97**	$\ll$	**17165.74**		**17001.48**	$\gg$	**16601.15**	$\ll$	**18601.02**	

(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	4		4		4		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	1266.62	$>$	1217.36	$\approx$	1251.36		1272.21	$>$	1217.36	$\approx$	1270.79	
BOS11–9	52.34	$\approx$	53.6	$<$	61.92		55.33	$\approx$	56.16	$\ll$	85.57	
BOS11–14	82.95	$\approx$	79.69	$\ll$	90.37		85.96	$\approx$	82.73	$\ll$	117.35	
BOS11–15	39.65	$\approx$	39.74	$\ll$	53.06		42.65	$\approx$	42.9	$\ll$	80.05	
BOS11–16	179.08	$\approx$	184.91	$\approx$	186.97		181.24	$\approx$	185.93	$\approx$	209.56	
BOS11–17	191.8	$\approx$	190.05	$\approx$	189.58		195.82	$\approx$	193.16	$\approx$	209.79	
BOS11–26	346.61	$\approx$	334.88	$\approx$	355.78		348.45	$\approx$	336.72	$\ll$	378.67	
BOS11–27	213.5	$\approx$	221.89	$\approx$	224.59		215	$\approx$	223.39	$\approx$	235.7	
BOS11–30	110.41	$\gg$	100.79	$\approx$	112.27		113.31	$\gg$	107.3	$\ll$	137.14	
BOS11–31	182.25	$\approx$	186.88	$\approx$	197.57		183.75	$\approx$	188.38	$<$	212.21	
BK12–28	935.25	$\approx$	944.39	$\approx$	913.55		937.04	$\approx$	946.17	$\approx$	921.25	
D05–18	364.67	$\approx$	363.43	$\approx$	366.23		366.54	$\approx$	365.3	$<$	389.12	
D05–19	440.24	$\approx$	433.06	$\approx$	437.46		448.35	$\approx$	436.22	$<$	468.87	
DF11–6	672.26	$\approx$	619.89	$\approx$	594.3		672.26	$\approx$	621.7	$\approx$	594.88	
DF11–7	1162.27	$\approx$	1211.18	$\gg$	1107.1		1164.23	$\approx$	1211.18	$\gg$	1107.1	
DF11–8	1037.38	$\approx$	1024.9	$>$	942.13		1039.3	$\approx$	1025.31	$>$	945.91	
DF11–22	1465.66	$\approx$	1451.97	$\approx$	1393.15		1467.55	$\approx$	1453.87	$\approx$	1409.12	
DF11–23	1242.36	$\approx$	1255.79	$\approx$	1201.41		1244.18	$\approx$	1255.79	$\approx$	1206.34	
DF11–24	1266.48	$\approx$	1253.11	$\approx$	1226.17		1268.37	$\approx$	1253.11	$\approx$	1233.69	
DF15–4	299.85	$>$	270.13	$\approx$	272.31		306.35	$>$	271.7	$\ll$	305.5	
DF15–5	1390.3	$\approx$	1349.58	$>$	1265.55		1392.77	$\approx$	1352.38	$>$	1283.51	
DF15–20	1786.19	$\approx$	1792.19	$>$	1724.15		1786.19	$\approx$	1792.19	$>$	1724.15	
DF15–21	2403.79	$\approx$	2403.31	$\gg$	2310.93		2403.79	$\approx$	2403.31	$\gg$	2310.93	
DF15–33	4039.13	$>$	3901.56	$>$	3804.27		4041.69	$>$	3902.53	$\approx$	3818.02	
DF15–35	978.44	$\approx$	952.81	$\approx$	958.36		980.23	$\approx$	954.6	$\approx$	976.03	
DRFN08–10	308.43	$\approx$	306.9	$\approx$	313.42		310.09	$\approx$	308.57	$<$	330.07	
DRFN08–11	436.15	$\approx$	432.7	$\approx$	419.05		439.24	$\approx$	434.24	$\approx$	436.69	
DO09–32	2784.68	$\approx$	2745.23	$\approx$	2702.02		2784.68	$\approx$	2745.23	$\approx$	2702.02	
FY17–25	504.26	$\approx$	486.46	$\approx$	496.45		508.47	$\approx$	486.46	$<$	514.21	
FRD12–29	748.15	$\approx$	750.98	$\approx$	745.6		750.09	$\approx$	752.91	$\approx$	763.71	
KS13–12	3057.93	$\approx$	3175.78	$\gg$	2929.77		3060.29	$\approx$	3177.98	$\gg$	2932.51	
STS13–13	1500.42	$<$	1590.96	$\gg$	1457.03		1502.43	$<$	1590.96	$\gg$	1462.8	
Aoyagi and Frechette (2009)	1266.62	$>$	1217.36	$\approx$	1251.36		1272.21	$>$	1217.36	$\approx$	1270.79	
Blonski et al. (2011)	1431.75	$\approx$	1426.73	$\ll$	1610.7		1471.58	$\approx$	1466.74	$\ll$	1966.5	
Bruttel and Kamecke (2012)	935.25	$\approx$	944.39	$\approx$	913.55		937.04	$\approx$	946.17	$\approx$	921.25	
Dal Bó (2005)	807.48	$\approx$	799.06	$\approx$	812.2		818.44	$\approx$	805.06	$\ll$	879.27	
Dal Bó and Fréchette (2011)	6868.86	$\approx$	6841.08	$\gg$	6580.33		6882.82	$\approx$	6847.87	$>$	6658.59	
Dal Bó and Fréchette (2015)	10921.32	$>$	10694.5	$\gg$	10457.73		10940.18	$>$	10705.87	$>$	10593.11	
Dreber et al. (2008)	746.97	$\approx$	742.4	$\approx$	743.19		752.84	$\approx$	746.31	$<$	787.77	
Duffy and Ochs (2009)	2784.68	$\approx$	2745.23	$\approx$	2702.02		2784.68	$\approx$	2745.23	$\approx$	2702.02	
Fréchette and Yuksel (2017)	504.26	$\approx$	486.46	$\approx$	496.45		508.47	$\approx$	486.46	$<$	514.21	
Fudenberg et al. (2012)	748.15	$\approx$	750.98	$\approx$	745.6		750.09	$\approx$	752.91	$\approx$	763.71	
Kagel and Schley (2013)	3057.93	$\approx$	3175.78	$\gg$	2929.77		3060.29	$\approx$	3177.98	$\gg$	2932.51	
Sherstyuk et al. (2013)	1500.42	$<$	1590.96	$\gg$	1457.03		1502.43	$<$	1590.96	$\gg$	1462.8	
Pooled	**31710.95**	$\approx$	**31564.59**	$>$	**31347.14**		**31863.45**	$\approx$	**31671.3**	$\ll$	**32546.75**	

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 7.

Table 11: Pure, mixed, or switching strategies? Best mixtures 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					
	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)	790.2	$>$	767.22	$\approx$	797.89		645.31	$\approx$	646.53	$\approx$	649.53	
Blonski et al. (2011)	674.3	$\gg$	626.71	$\ll$	843		713.8	$\gg$	670.62	$\ll$	875.74	
Bruttel and Kamecke (2012)	590.68	$\approx$	581	$\approx$	592.44		585.42	$\approx$	570.56	$\approx$	570.87	
Dal Bó (2005)	394.44	$\gg$	365.35	$\ll$	393		407.86	$\gg$	378.95	$<$	404.78	
Dal Bó and Fréchette (2011)	3847.44	$\approx$	3775.96	$\approx$	3802.38		3536.73	$\approx$	3589.36	$\approx$	3524.77	
Dal Bó and Fréchette (2015)	5522.08	$\gg$	5280.06	$\approx$	5318.33		5259.64	$\gg$	5037.82	$\approx$	5057.54	
Dreber et al. (2008)	459.11	$\approx$	462.48	$\approx$	481.64		478.09	$\approx$	466.13	$\approx$	482.86	
Duffy and Ochs (2009)	1128.03	$\approx$	1111.23	$\approx$	1123.74		1047.59	$\approx$	1053.04	$\approx$	1049.79	
Fréchette and Yuksel (2017)	181.98	$\gg$	161.75	$<$	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)	1813.04	$\approx$	1862.88	$\gg$	1781.74		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	**16820.09**	$\gg$	**16417.97**	$\ll$	**17165.74**		**16077.95**	$>$	**15853.82**	$\ll$	**16335.33**	
Second halves per session												
Aoyagi and Frechette (2009)	478.46	$>$	451.68	$\approx$	480.65		363.58	$\approx$	368.23	$\approx$	368.89	
Blonski et al. (2011)	965.51	$\gg$	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)	464.42	$\approx$	446.63	$\ll$	474.71		475.11	$\approx$	456.4	$\approx$	469.98	
Dal Bó and Fréchette (2011)	3052.84	$\approx$	3081.92	$\approx$	3092.35		2737.11	$<$	2875.64	$>$	2721.88	
Dal Bó and Fréchette (2015)	5571.71	$>$	5419.19	$\approx$	5478.11		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)	1628.02	$\approx$	1628.89	$\approx$	1609.73		1381.01	$\approx$	1416.71	$\approx$	1392.49	
Fréchette and Yuksel (2017)	356.12	$\approx$	332.38	$<$	344.23		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)	1213.26	$<$	1301.76	$\gg$	1203.05		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	**15523.37**	$\approx$	**15443.11**	$\ll$	**16124.95**		**14387.48**	$<$	**14656.93**	$\approx$	**14961.61**	

Note: Relation signs are used as defined above (Table 7). "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 12: Table 11 by treatments – Pure, mixed, or switching strategies? Best mixtures without Semi-Grim

(a) First halves per session

	Best mixture of pure strategies						Best mixture of generalized pure strategies					
	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	
AF09–34	790.2	$>$	767.22	$\approx$	797.89		645.31	$\approx$	646.53	$\approx$	649.53	
BOS11–9	45.36	$\approx$	43.78	$\ll$	55.54		39.68	$\approx$	34.86	$\ll$	47.44	
BOS11–14	61.42	$\approx$	59.77	$\ll$	72.93		52.96	$\approx$	51.25	$\ll$	69.19	
BOS11–15	29.87	$\approx$	29.93	$\ll$	43.24		31.14	$\gg$	22.09	$\ll$	45.44	
BOS11–16	106.92	$\approx$	103.63	$\ll$	118.03		106.93	$\approx$	109.67	$\approx$	112.68	
BOS11–17	48.66	$<$	53.62	$\approx$	57.44		56.31	$\approx$	57.46	$\approx$	60.52	
BOS11–26	162.85	$\gg$	140.64	$\ll$	153.41		162.52	$>$	139.66	$<$	158.95	
BOS11–27	65.77	$>$	52.93	$\ll$	65.67		64.41	$\gg$	54.54	$<$	70.16	
BOS11–30	30.38	$\approx$	30.39	$\ll$	43.76		42.87	$\approx$	42.83	$\ll$	56.32	
BOS11–31	88.41	$>$	79.66	$\ll$	103.59		88.38	$\approx$	89.66	$\approx$	94.45	
BK12–28	590.68	$\approx$	581	$\approx$	592.44		585.42	$\approx$	570.56	$\approx$	570.87	
D05–18	153.77	$\gg$	136.89	$\ll$	151.01		154.75	$\gg$	138.83	$\ll$	151.81	
D05–19	238.55	$\approx$	226.34	$\approx$	233.48		248.85	$\approx$	235.87	$\approx$	242.34	
DF11–6	379.95	$>$	357.02	$\approx$	347.58		339.47	$\approx$	347.55	$\approx$	316.5	
DF11–7	549.14	$>$	513.5	$\approx$	514.91		526.77	$\approx$	523.18	$>$	492.09	
DF11–8	581.17	$\approx$	568.86	$\approx$	552.52		563.75	$\approx$	552.71	$\approx$	523.57	
DF11–22	758.46	$\approx$	738.85	$\approx$	742.48		707.53	$\approx$	697.75	$\approx$	694.38	
DF11–23	715.77	$\approx$	735.67	$\approx$	706.54		619.29	$<$	667.99	$\approx$	656.5	
DF11–24	843.2	$\approx$	838.63	$\approx$	830.64		740.3	$\approx$	758.9	$\approx$	720.58	
DF15–4	170.82	$\approx$	158.62	$\approx$	164.65		175.94	$\approx$	180.54	$>$	168.75	
DF15–5	761.26	$\approx$	724.76	$\approx$	696.32		737.26	$\approx$	726.27	$>$	674.63	
DF15–20	1003.49	$>$	955.28	$\approx$	958.24		960.95	$>$	913.96	$\approx$	900.5	
DF15–21	1334.08	$\approx$	1316.87	$\gg$	1274.95		1234.63	$\approx$	1210.32	$\approx$	1218.06	
DF15–33	1900.51	$\gg$	1796.11	$\approx$	1810.13		1800.02	$\gg$	1671.27	$\approx$	1691.77	
DF15–35	330.58	$>$	307.28	$\approx$	319.5		310.13	$\approx$	293.31	$\approx$	295.78	
DRFN08–10	197.04	$\approx$	195.4	$\approx$	208.99		208.83	$\approx$	197.36	$\approx$	213.97	
DRFN08–11	259.69	$\approx$	264.28	$\approx$	264.24		264.48	$\approx$	263.16	$\approx$	258.39	
DO09–32	1128.03	$\approx$	1111.23	$\approx$	1123.74		1047.59	$\approx$	1053.04	$\approx$	1049.79	
FY17–25	181.98	$\gg$	161.75	$<$	176.5		188.5	$\approx$	183.48	$\approx$	175.59	
FRD12–29	356.73	$>$	331.44	$<$	347.07		319.45	$\approx$	308.6	$\approx$	320.55	
KS13–12	1813.04	$\approx$	1862.88	$\gg$	1781.74		1761.98	$\approx$	1780.97	$>$	1694.94	
STS13–13	926.9	$\approx$	953.91	$>$	912.67		865.67	$\approx$	907.14	$>$	858.65	
Aoyagi and Frechette (2009)	790.2	$>$	767.22	$\approx$	797.89		645.31	$\approx$	646.53	$\approx$	649.53	
Blonski et al. (2011)	674.3	$\gg$	626.71	$\ll$	843		713.8	$\gg$	670.62	$\ll$	875.74	
Bruttel and Kamecke (2012)	590.68	$\approx$	581	$\approx$	592.44		585.42	$\approx$	570.56	$\approx$	570.87	
Dal Bó (2005)	394.44	$\gg$	365.35	$\ll$	393		407.86	$\gg$	378.95	$<$	404.78	
Dal Bó and Fréchette (2011)	3847.44	$\approx$	3775.96	$\approx$	3802.38		3536.73	$\approx$	3589.36	$\approx$	3524.77	
Dal Bó and Fréchette (2015)	5522.08	$\gg$	5280.06	$\approx$	5318.33		5259.64	$\gg$	5037.82	$\approx$	5057.54	
Dreber et al. (2008)	459.11	$\approx$	462.48	$\approx$	481.64		478.09	$\approx$	466.13	$\approx$	482.86	
Duffy and Ochs (2009)	1128.03	$\approx$	1111.23	$\approx$	1123.74		1047.59	$\approx$	1053.04	$\approx$	1049.79	
Fréchette and Yuksel (2017)	181.98	$\gg$	161.75	$<$	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)	1813.04	$\approx$	1862.88	$\gg$	1781.74		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	**16820.09**	$\gg$	**16417.97**	$\ll$	**17165.74**		**16077.95**	$>$	**15853.82**	$\ll$	**16335.33**	

(b) Second halves per session

	Best mixture of pure strategies						Best mixture of generalized pure strategies					
	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	
AF09–34	1266.62	$>$	1217.36	$\approx$	1251.36		998.99	$\approx$	1011.31	$\approx$	989.06	
BOS11–9	52.34	$\approx$	53.6	$<$	61.92		56.83	$<$	67.33	$\approx$	63.72	
BOS11–14	82.95	$\approx$	79.69	$\ll$	90.37		75.9	$\approx$	70.2	$<$	77.47	
BOS11–15	39.65	$\approx$	39.74	$\ll$	53.06		43.38	$\gg$	34.25	$\ll$	55.25	
BOS11–16	179.08	$\approx$	184.91	$\approx$	186.97		170.31	$\approx$	187.28	$\approx$	176.88	
BOS11–17	191.8	$\approx$	190.05	$\approx$	189.58		191.55	$\approx$	192.15	$\approx$	194.05	
BOS11–26	346.61	$\approx$	334.88	$\approx$	355.78		328.04	$\approx$	330.67	$\approx$	340.16	
BOS11–27	213.5	$\approx$	221.89	$\approx$	224.59		214.12	$\approx$	222.33	$\approx$	218.59	
BOS11–30	110.41	$\gg$	100.79	$\approx$	112.27		116.18	$\gg$	109.21	$\approx$	116.12	
BOS11–31	182.25	$\approx$	186.88	$\approx$	197.57		183.27	$\approx$	185.32	$\approx$	188.11	
BK12–28	935.25	$\approx$	944.39	$\approx$	913.55		915.48	$\approx$	926.52	$>$	879.03	
D05–18	364.67	$\approx$	363.43	$\approx$	366.23		360.73	$\approx$	361.89	$\approx$	355.8	
D05–19	440.24	$\approx$	433.06	$\approx$	437.46		442.98	$\approx$	435.7	$\approx$	431.03	
DF11–6	672.26	$\approx$	619.89	$\approx$	594.3		584.49	$\approx$	615.02	$\approx$	531.24	
DF11–7	1162.27	$\approx$	1211.18	$\gg$	1107.1		1135.74	$\approx$	1179.37	$\gg$	996.23	
DF11–8	1037.38	$\approx$	1024.9	$>$	942.13		971.84	$\approx$	971.32	$>$	874.03	
DF11–22	1465.66	$\approx$	1451.97	$\approx$	1393.15		1348.86	$\approx$	1333.85	$>$	1269.49	
DF11–23	1242.36	$\approx$	1255.79	$\approx$	1201.41		1105.85	$\approx$	1135.28	$>$	1043.89	
DF11–24	1266.48	$\approx$	1253.11	$\approx$	1226.17		1076.38	$\approx$	1105.99	$\approx$	1038.16	
DF15–4	299.85	$>$	270.13	$\approx$	272.31		298.69	$\approx$	283.7	$\approx$	279.23	
DF15–5	1390.3	$\approx$	1349.58	$>$	1265.55		1343.25	$\approx$	1361.6	$\gg$	1240.17	
DF15–20	1786.19	$\approx$	1792.19	$>$	1724.15		1696.3	$\approx$	1689.33	$\gg$	1589.04	
DF15–21	2403.79	$\approx$	2403.31	$\gg$	2310.93		2152.39	$\approx$	2212.29	$\gg$	2054.9	
DF15–33	4039.13	$>$	3901.56	$>$	3804.27		3731.12	$>$	3591.79	$\approx$	3531.59	
DF15–35	978.44	$\approx$	952.81	$\approx$	958.36		910.07	$>$	854.66	$\approx$	870.49	
DRFN08–10	308.43	$\approx$	306.9	$\approx$	313.42		310.63	$\approx$	308.06	$\approx$	311.09	
DRFN08–11	436.15	$\approx$	432.7	$\approx$	419.05		419.57	$\approx$	427.47	$>$	404.08	
DO09–32	2784.68	$\approx$	2745.23	$\approx$	2702.02		2401.04	$\approx$	2451.43	$\gg$	2321.8	
FY17–25	504.26	$\approx$	486.46	$\approx$	496.45		475.52	$\approx$	454.57	$\approx$	459.64	
FRD12–29	748.15	$\approx$	750.98	$\approx$	745.6		636.43	$<$	685.17	$>$	644.13	
KS13–12	3057.93	$\approx$	3175.78	$\gg$	2929.77		2935.17	$\approx$	3003.32	$\gg$	2783.52	
STS13–13	1500.42	$<$	1590.96	$\gg$	1457.03		1358.57	$<$	1493.91	$\gg$	1325.94	
Aoyagi and Frechette (2009)	1266.62	$>$	1217.36	$\approx$	1251.36		998.99	$\approx$	1011.31	$\approx$	989.06	
Blonski et al. (2011)	1431.75	$\approx$	1426.73	$\ll$	1610.7		1450.51	$\approx$	1471.96	$\ll$	1608.54	
Bruttel and Kamecke (2012)	935.25	$\approx$	944.39	$\approx$	913.55		915.48	$\approx$	926.52	$>$	879.03	
Dal Bó (2005)	807.48	$\approx$	799.06	$\approx$	812.2		807.96	$\approx$	803.27	$\approx$	797.46	
Dal Bó and Fréchette (2011)	6868.86	$\approx$	6841.08	$\gg$	6580.33		6260.83	$\approx$	6383.92	$\gg$	5874.18	
Dal Bó and Fréchette (2015)	10921.32	$>$	10694.5	$\gg$	10457.73		10177.65	$\approx$	10037.43	$\gg$	9693.72	
Dreber et al. (2008)	746.97	$\approx$	742.4	$\approx$	743.19		734.98	$\approx$	740.32	$\approx$	728.28	
Duffy and Ochs (2009)	2784.68	$\approx$	2745.23	$\approx$	2702.02		2401.04	$\approx$	2451.43	$\gg$	2321.8	
Fréchette and Yuksel (2017)	504.26	$\approx$	486.46	$\approx$	496.45		475.52	$\approx$	454.57	$\approx$	459.64	
Fudenberg et al. (2012)	748.15	$\approx$	750.98	$\approx$	745.6		636.43	$<$	685.17	$>$	644.13	
Kagel and Schley (2013)	3057.93	$\approx$	3175.78	$\gg$	2929.77		2935.17	$\approx$	3003.32	$\gg$	2783.52	
Sherstyuk et al. (2013)	1500.42	$<$	1590.96	$\gg$	1457.03		1358.57	$<$	1493.91	$\gg$	1325.94	
Pooled	**31710.95**	$\approx$	**31564.59**	$>$	**31347.14**		**29421.33**	$\approx$	**29743.95**	$\gg$	**28851.02**	

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 7.

Table 13: 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 + SG2	
Specification												
# Models evaluated	4		4		4				1		1	
# Pars estimated (by treatment)	16		16		82				3		3	
# Parameters accounted for	3–5		3–5		12–35				3		3	
First halves per session												
Aoyagi and Frechette (2009)	790.2	$>$	767.22	$\approx$	797.89		767.22	$\gg$	694.72	$\approx$	739.24	
Blonski et al. (2011)	674.3	$\gg$	626.71	$\ll$	843		626.71	$\gg$	549.45	$\ll$	621.23	
Bruttel and Kamecke (2012)	590.68	$\approx$	581	$\approx$	592.44		581	$\approx$	567.86	$\approx$	566	
Dal Bó (2005)	394.44	$\gg$	365.35	$\ll$	393		365.35	$\approx$	358.51	$<$	374.94	
Dal Bó and Fréchette (2011)	3847.44	$\approx$	3775.96	$\approx$	3802.38		3775.96	$\gg$	3533.99	$\approx$	3550.1	
Dal Bó and Fréchette (2015)	5522.08	$\gg$	5280.06	$\approx$	5318.33		5280.06	$\gg$	4991.74	$\approx$	5014.35	
Dreber et al. (2008)	459.11	$\approx$	462.48	$\approx$	481.64		462.48	$>$	437.17	$\approx$	449.12	
Duffy and Ochs (2009)	1128.03	$\approx$	1111.23	$\approx$	1123.74		1111.23	$\approx$	1090.22	$\approx$	1086.17	
Fréchette and Yuksel (2017)	181.98	$\gg$	161.75	$<$	176.5		161.75	$\approx$	161.45	$\approx$	162.3	
Fudenberg et al. (2012)	356.73	$>$	331.44	$<$	347.07		331.44	$>$	291.43	$\approx$	295.84	
Kagel and Schley (2013)	1813.04	$\approx$	1862.88	$\gg$	1781.74		1862.88	$>$	1782.82	$>$	1706.03	
Sherstyuk et al. (2013)	926.9	$\approx$	953.91	$>$	912.67		953.91	$\approx$	912.8	$\approx$	899.24	
Pooled	**16820.09**	$\gg$	**16417.97**	$\ll$	**17165.74**		**16417.97**	$\gg$	**15481.59**	$\approx$	**15573.98**	
Second halves per session												
Aoyagi and Frechette (2009)	478.46	$>$	451.68	$\approx$	480.65		478.46	$\gg$	389.24	$<$	429.54	
Blonski et al. (2011)	965.51	$\gg$	914.28	$\ll$	1140.5		965.51	$>$	867.87	$\approx$	892.86	
Bruttel and Kamecke (2012)	342.17	$\approx$	361.38	$\approx$	348.88		342.17	$\approx$	347.4	$\approx$	345.74	
Dal Bó (2005)	464.42	$\approx$	446.63	$\ll$	474.71		464.42	$>$	424.44	$\approx$	430.13	
Dal Bó and Fréchette (2011)	3052.84	$\approx$	3081.92	$\approx$	3092.35		3052.84	$>$	2817.31	$\approx$	2764.96	
Dal Bó and Fréchette (2015)	5571.71	$>$	5419.19	$\approx$	5478.11		5571.71	$\gg$	5043.81	$\approx$	5118.03	
Dreber et al. (2008)	287.58	$\approx$	285.34	$<$	303.79		287.58	$\approx$	264.94	$<$	288.98	
Duffy and Ochs (2009)	1628.02	$\approx$	1628.89	$\approx$	1609.73		1628.02	$\gg$	1403.03	$\approx$	1444.97	
Fréchette and Yuksel (2017)	356.12	$\approx$	332.38	$<$	344.23		356.12	$>$	313.5	$\approx$	319.17	
Fudenberg et al. (2012)	443.13	$\approx$	439.28	$\approx$	444.41		443.13	$>$	380.75	$\approx$	370.44	
Kagel and Schley (2013)	1213.26	$<$	1301.76	$\gg$	1203.05		1213.26	$\approx$	1211.37	$>$	1139.73	
Sherstyuk et al. (2013)	587.45	$<$	640.1	$>$	597.28		587.45	$\approx$	586.72	$\approx$	562.71	
Pooled	**15523.37**	$\approx$	**15443.11**	$\ll$	**16124.95**		**15523.37**	$\gg$	**14159.8**	$\approx$	**14216.67**	

Note: This table extends Table 9 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.

Table 14: Table 13 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 + SG2
Specification										
# Models evaluated	4		4		4			1		1
# Pars estimated (by treatment)	16		16		82			3		3
# Parameters accounted for	3–5		3–5		12–35			3		3
AF09–34	790.2	$>$	767.22	$\approx$	797.89	767.22	$\gg$	694.72	$\approx$	739.24
BOS11–9	45.36	$\approx$	43.78	$\ll$	55.54	43.78	$\gg$	28.34	$\approx$	29.99
BOS11–14	61.42	$\approx$	59.77	$\ll$	72.93	59.77	$\gg$	43.74	$\ll$	56.52
BOS11–15	29.87	$\approx$	29.93	$\ll$	43.24	29.93	$\gg$	14.27	$\ll$	27.48
BOS11–16	106.92	$\approx$	103.63	$\ll$	118.03	103.63	$\approx$	102.31	$\approx$	102.4
BOS11–17	48.66	$<$	53.62	$\approx$	57.44	53.62	$\approx$	48.33	$\approx$	53.95
BOS11–26	162.85	$\gg$	140.64	$\ll$	153.41	140.64	$\approx$	130.46	$\ll$	146.7
BOS11–27	65.77	$>$	52.93	$\ll$	65.67	52.93	$\approx$	49.15	$\ll$	57.95
BOS11–30	30.38	$\approx$	30.39	$\ll$	43.76	30.39	$\gg$	15.97	$\ll$	30.27
BOS11–31	88.41	$>$	79.66	$\ll$	103.59	79.66	$\approx$	86.84	$\approx$	85.93
BK12–28	590.68	$\approx$	581	$\approx$	592.44	581	$\approx$	567.86	$\approx$	566
D05–18	153.77	$\gg$	136.89	$\ll$	151.01	136.89	$\approx$	136.21	$<$	141.56
D05–19	238.55	$\approx$	226.34	$\approx$	233.48	226.34	$\approx$	220.17	$<$	231.25
DF11–6	379.95	$>$	357.02	$\approx$	347.58	357.02	$>$	333.68	$\approx$	341.04
DF11–7	549.14	$>$	513.5	$\approx$	514.91	513.5	$\approx$	512.99	$\approx$	525.32
DF11–8	581.17	$\approx$	568.86	$\approx$	552.52	568.86	$\approx$	551.81	$\approx$	549.33
DF11–22	758.46	$\approx$	738.85	$\approx$	742.48	738.85	$\approx$	684.04	$\approx$	689.12
DF11–23	715.77	$\approx$	735.67	$\approx$	706.54	735.67	$\gg$	656.29	$\approx$	643.12
DF11–24	843.2	$\approx$	838.63	$\approx$	830.64	838.63	$>$	779.03	$\approx$	786.01
DF15–4	170.82	$\approx$	158.62	$\approx$	164.65	158.62	$\approx$	149.71	$\ll$	174.17
DF15–5	761.26	$\approx$	724.76	$\approx$	696.32	724.76	$\approx$	716.51	$\approx$	689.89
DF15–20	1003.49	$>$	955.28	$\approx$	958.24	955.28	$\approx$	914.2	$\approx$	931.26
DF15–21	1334.08	$\approx$	1316.87	$\gg$	1274.95	1316.87	$\gg$	1196.29	$\approx$	1181.18
DF15–33	1900.51	$\gg$	1796.11	$\approx$	1810.13	1796.11	$>$	1714.32	$\approx$	1735.23
DF15–35	330.58	$>$	307.28	$\approx$	319.5	307.28	$\approx$	283.21	$\approx$	285.12
DRFN08–10	197.04	$\approx$	195.4	$\approx$	208.99	195.4	$\approx$	180.96	$\approx$	190.57
DRFN08–11	259.69	$\approx$	264.28	$\approx$	264.24	264.28	$\approx$	254.11	$\approx$	256.45
DO09–32	1128.03	$\approx$	1111.23	$\approx$	1123.74	1111.23	$\approx$	1090.22	$\approx$	1086.17
FY17–25	181.98	$\gg$	161.75	$<$	176.5	161.75	$\approx$	161.45	$\approx$	162.3
FRD12–29	356.73	$>$	331.44	$<$	347.07	331.44	$>$	291.43	$\approx$	295.84
KS13–12	1813.04	$\approx$	1862.88	$\gg$	1781.74	1862.88	$>$	1782.82	$>$	1706.03
STS13–13	926.9	$\approx$	953.91	$>$	912.67	953.91	$\approx$	912.8	$\approx$	899.24
Aoyagi and Frechette (2009)	790.2	$>$	767.22	$\approx$	797.89	767.22	$\gg$	694.72	$\approx$	739.24
Blonski et al. (2011)	674.3	$\gg$	626.71	$\ll$	843	626.71	$\gg$	549.45	$\ll$	621.23
Bruttel and Kamecke (2012)	590.68	$\approx$	581	$\approx$	592.44	581	$\approx$	567.86	$\approx$	566
Dal Bó (2005)	394.44	$\gg$	365.35	$\ll$	393	365.35	$\approx$	358.51	$<$	374.94
Dal Bó and Fréchette (2011)	3847.44	$\approx$	3775.96	$\approx$	3802.38	3775.96	$\gg$	3533.99	$\approx$	3550.1
Dal Bó and Fréchette (2015)	5522.08	$\gg$	5280.06	$\approx$	5318.33	5280.06	$\gg$	4991.74	$\approx$	5014.35
Dreber et al. (2008)	459.11	$\approx$	462.48	$\approx$	481.64	462.48	$>$	437.17	$\approx$	449.12
Duffy and Ochs (2009)	1128.03	$\approx$	1111.23	$\approx$	1123.74	1111.23	$\approx$	1090.22	$\approx$	1086.17
Fréchette and Yuksel (2017)	181.98	$\gg$	161.75	$<$	176.5	161.75	$\approx$	161.45	$\approx$	162.3
Fudenberg et al. (2012)	356.73	$>$	331.44	$<$	347.07	331.44	$>$	291.43	$\approx$	295.84
Kagel and Schley (2013)	1813.04	$\approx$	1862.88	$\gg$	1781.74	1862.88	$>$	1782.82	$>$	1706.03
Sherstyuk et al. (2013)	926.9	$\approx$	953.91	$>$	912.67	953.91	$\approx$	912.8	$\approx$	899.24
Pooled	**16820.09**	$\gg$	**16417.97**	$\ll$	**17165.74**	**16417.97**	$\gg$	**15481.59**	$\approx$	**15573.98**

(b) Second halves per session

	Best mixture of pure strategies									
	No Switching		Random Switching		Markov Switching	Best Switching		Semi-Grim		AD + SG2
Specification										
# Models evaluated	4		4		4			1		1
# Pars estimated (by treatment)	16		16		82			3		3
# Parameters accounted for	3–5		3–5		12–35			3		3
AF09–34	1266.62	$>$	1217.36	$\approx$	1251.36	1251.36	$\gg$	1085.04	$<$	1178.07
BOS11–9	52.34	$\approx$	53.6	$<$	61.92	61.92	$>$	50.83	$\approx$	51.64
BOS11–14	82.95	$\approx$	79.69	$\ll$	90.37	90.37	$\gg$	61.74	$\ll$	69.39
BOS11–15	39.65	$\approx$	39.74	$\ll$	53.06	53.06	$\gg$	21.88	$\ll$	37.94
BOS11–16	179.08	$\approx$	184.91	$\approx$	186.97	186.97	$\approx$	178.84	$\approx$	174.16
BOS11–17	191.8	$\approx$	190.05	$\approx$	189.58	189.58	$\approx$	183.03	$\approx$	181.55
BOS11–26	346.61	$\approx$	334.88	$\approx$	355.78	355.78	$\approx$	317.33	$\approx$	317.2
BOS11–27	213.5	$\approx$	221.89	$\approx$	224.59	224.59	$\approx$	231.72	$\approx$	215.78
BOS11–30	110.41	$\gg$	100.79	$\approx$	112.27	112.27	$\gg$	93.22	$\approx$	99.81
BOS11–31	182.25	$\approx$	186.88	$\approx$	197.57	197.57	$\approx$	200.87	$\approx$	188.19
BK12–28	935.25	$\approx$	944.39	$\approx$	913.55	913.55	$\approx$	916.81	$\approx$	903.59
D05–18	364.67	$\approx$	363.43	$\approx$	366.23	366.23	$\approx$	351.89	$\approx$	353.01
D05–19	440.24	$\approx$	433.06	$\approx$	437.46	437.46	$\approx$	418.28	$\approx$	423.88
DF11–6	672.26	$\approx$	619.89	$\approx$	594.3	594.3	$\approx$	599.29	$\approx$	579.61
DF11–7	1162.27	$\approx$	1211.18	$\gg$	1107.1	1107.1	$<$	1199.27	$\approx$	1168.14
DF11–8	1037.38	$\approx$	1024.9	$>$	942.13	942.13	$\approx$	985.67	$\approx$	956.71
DF11–22	1465.66	$\approx$	1451.97	$\approx$	1393.15	1393.15	$\approx$	1331.85	$\approx$	1303.37
DF11–23	1242.36	$\approx$	1255.79	$\approx$	1201.41	1201.41	$>$	1113.94	$\approx$	1113.42
DF11–24	1266.48	$\approx$	1253.11	$\approx$	1226.17	1226.17	$>$	1122.26	$\approx$	1162.51
DF15–4	299.85	$>$	270.13	$\approx$	272.31	272.31	$\approx$	260.02	$<$	283.22
DF15–5	1390.3	$\approx$	1349.58	$>$	1265.55	1265.55	$\approx$	1331.49	$\approx$	1291.93
DF15–20	1786.19	$\approx$	1792.19	$>$	1724.15	1724.15	$\approx$	1716.02	$\approx$	1702.32
DF15–21	2403.79	$\approx$	2403.31	$\gg$	2310.93	2310.93	$\approx$	2213.88	$\approx$	2176.02
DF15–33	4039.13	$>$	3901.56	$>$	3804.27	3804.27	$\approx$	3678.61	$\approx$	3654.5
DF15–35	978.44	$\approx$	952.81	$\approx$	958.36	958.36	$\gg$	856.78	$\approx$	842.09
DRFN08–10	308.43	$\approx$	306.9	$\approx$	313.42	313.42	$>$	270.35	$<$	297.4
DRFN08–11	436.15	$\approx$	432.7	$\approx$	419.05	419.05	$\approx$	425.89	$\approx$	423.94
DO09–32	2784.68	$\approx$	2745.23	$\approx$	2702.02	2702.02	$\gg$	2526.02	$\approx$	2569.75
FY17–25	504.26	$\approx$	486.46	$\approx$	496.45	496.45	$\approx$	473.84	$\approx$	477.72
FRD12–29	748.15	$\approx$	750.98	$\approx$	745.6	745.6	$>$	668.18	$\approx$	652.38
KS13–12	3057.93	$\approx$	3175.78	$\gg$	2929.77	2929.77	$\approx$	3009.28	$>$	2831.74
STS13–13	1500.42	$<$	1590.96	$\gg$	1457.03	1457.03	$\approx$	1507.33	$\approx$	1459.95
Aoyagi and Frechette (2009)	1266.62	$>$	1217.36	$\approx$	1251.36	1251.36	$\gg$	1085.04	$<$	1178.07
Blonski et al. (2011)	1431.75	$\approx$	1426.73	$\ll$	1610.7	1610.7	$\gg$	1369.51	$\approx$	1365.7
Bruttel and Kamecke (2012)	935.25	$\approx$	944.39	$\approx$	913.55	913.55	$\approx$	916.81	$\approx$	903.59
Dal Bó (2005)	807.48	$\approx$	799.06	$\approx$	812.2	812.2	$\approx$	772.3	$\approx$	779.01
Dal Bó and Fréchette (2011)	6868.86	$\approx$	6841.08	$\gg$	6580.33	6580.33	$\approx$	6368.43	$\approx$	6299.92
Dal Bó and Fréchette (2015)	10921.32	$>$	10694.5	$\gg$	10457.73	10457.73	$\gg$	10074.29	$\approx$	9967.59
Dreber et al. (2008)	746.97	$\approx$	742.4	$\approx$	743.19	743.19	$\approx$	698.35	$\approx$	723.44
Duffy and Ochs (2009)	2784.68	$\approx$	2745.23	$\approx$	2702.02	2702.02	$\gg$	2526.02	$\approx$	2569.75
Fréchette and Yuksel (2017)	504.26	$\approx$	486.46	$\approx$	496.45	496.45	$\approx$	473.84	$\approx$	477.72
Fudenberg et al. (2012)	748.15	$\approx$	750.98	$\approx$	745.6	745.6	$>$	668.18	$\approx$	652.38
Kagel and Schley (2013)	3057.93	$\approx$	3175.78	$\gg$	2929.77	2929.77	$\approx$	3009.28	$>$	2831.74
Sherstyuk et al. (2013)	1500.42	$<$	1590.96	$\gg$	1457.03	1457.03	$\approx$	1507.33	$\approx$	1459.95
Pooled	**31710.95**	$\approx$	**31564.59**	$>$	**31347.14**	**31347.14**	$\gg$	**29578.81**	$>$	**29318.3**

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 7.

Table 15: 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											
	No Switching		Random Switching		Markov Switching		Best Switching		Semi-Grim		AD + SG2	
Specification												
# Models evaluated	4		4		4				1		1	
# Pars estimated (by treatment)	32		32		98				3		3	
# Parameters accounted for	6–10		6–10		15-35				3		3	
First halves per session												
Aoyagi and Frechette (2009)	645.31	$\approx$	646.53	$\approx$	649.53		646.53	$\approx$	694.72	$\approx$	739.24	
Blonski et al. (2011)	713.8	$\gg$	670.62	$\ll$	875.74		670.62	$\gg$	549.45	$\ll$	621.23	
Bruttel and Kamecke (2012)	585.42	$\approx$	570.56	$\approx$	570.87		570.56	$\approx$	567.86	$\approx$	566	
Dal Bó (2005)	407.86	$\gg$	378.95	$<$	404.78		378.95	$>$	358.51	$<$	374.94	
Dal Bó and Fréchette (2011)	3536.73	$\approx$	3589.36	$\approx$	3524.77		3589.36	$\approx$	3533.99	$\approx$	3550.1	
Dal Bó and Fréchette (2015)	5259.64	$\gg$	5037.82	$\approx$	5057.54		5037.82	$\approx$	4991.74	$\approx$	5014.35	
Dreber et al. (2008)	478.09	$\approx$	466.13	$\approx$	482.86		466.13	$\gg$	437.17	$\approx$	449.12	
Duffy and Ochs (2009)	1047.59	$\approx$	1053.04	$\approx$	1049.79		1053.04	$\approx$	1090.22	$\approx$	1086.17	
Fréchette and Yuksel (2017)	188.5	$\approx$	183.48	$\approx$	175.59		183.48	$\gg$	161.45	$\approx$	162.3	
Fudenberg et al. (2012)	319.45	$\approx$	308.6	$\approx$	320.55		308.6	$\approx$	291.43	$\approx$	295.84	
Kagel and Schley (2013)	1761.98	$\approx$	1780.97	$>$	1694.94		1780.97	$\approx$	1782.82	$>$	1706.03	
Sherstyuk et al. (2013)	865.67	$\approx$	907.14	$>$	858.65		907.14	$\approx$	912.8	$\approx$	899.24	
Pooled	**16077.95**	$>$	**15853.82**	$\ll$	**16335.33**		**15853.82**	$\gg$	**15481.59**	$\approx$	**15573.98**	
Second halves per session												
Aoyagi and Frechette (2009)	363.58	$\approx$	368.23	$\approx$	368.89		363.58	$\approx$	389.24	$<$	429.54	
Blonski et al. (2011)	992.44	$\approx$	993.71	$\ll$	1154.56		992.44	$\gg$	867.87	$\approx$	892.86	
Bruttel and Kamecke (2012)	344.88	$\approx$	358.12	$\approx$	347.08		344.88	$\approx$	347.4	$\approx$	345.74	
Dal Bó (2005)	475.11	$\approx$	456.4	$\approx$	469.98		475.11	$\gg$	424.44	$\approx$	430.13	
Dal Bó and Fréchette (2011)	2737.11	$<$	2875.64	$>$	2721.88		2737.11	$\approx$	2817.31	$\approx$	2764.96	
Dal Bó and Fréchette (2015)	5164.78	$\approx$	5105.6	$\approx$	5116.42		5164.78	$\approx$	5043.81	$\approx$	5118.03	
Dreber et al. (2008)	295.06	$\approx$	297.88	$\approx$	303.03		295.06	$\approx$	264.94	$<$	288.98	
Duffy and Ochs (2009)	1381.01	$\approx$	1416.71	$\approx$	1392.49		1381.01	$\approx$	1403.03	$\approx$	1444.97	
Fréchette and Yuksel (2017)	309.63	$\approx$	304.7	$\approx$	308.78		309.63	$\approx$	313.5	$\approx$	319.17	
Fudenberg et al. (2012)	373.44	$\approx$	395.32	$\approx$	376.62		373.44	$\approx$	380.75	$\approx$	370.44	
Kagel and Schley (2013)	1170.12	$\approx$	1224.37	$>$	1143.67		1170.12	$\approx$	1211.37	$>$	1139.73	
Sherstyuk et al. (2013)	527.09	$\approx$	590.16	$\approx$	567.63		527.09	$\approx$	586.72	$\approx$	562.71	
Pooled	**14387.48**	$<$	**14656.93**	$\approx$	**14961.61**		**14387.48**	$\approx$	**14159.8**	$\approx$	**14216.67**	

Note: This table extends Table 11 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 16: Table 15 by treatments – Best mixtures of generalized strategies in relation to a Semi-Grim strategy

(a) First halves per session

	Best mixture of generalized strategies										Semi-Grim		AD + SG2
	No Switching		Random Switching		Markov Switching		Best Switching						
Specification													
# Models evaluated	4		4		4						1		1
# Pars estimated (by treatment)	32		32		98						3		3
# Parameters accounted for	6–10		6–10		15-35						3		3
AF09–34	645.31	≈	646.53	≈	649.53		646.53	≈	694.72	≈	739.24		
BOS11–9	39.68	≈	34.86	≪	47.44		34.86	≫	28.34	≈	29.99		
BOS11–14	52.96	≈	51.25	≪	69.19		51.25	>	43.74	≪	56.52		
BOS11–15	31.14	≫	22.09	≪	45.44		22.09	≫	14.27	≪	27.48		
BOS11–16	106.93	≈	109.67	≈	112.68		109.67	≈	102.31	≈	102.4		
BOS11–17	56.31	≈	57.46	≈	60.52		57.46	≫	48.33	≈	53.95		
BOS11–26	162.52	>	139.66	<	158.95		139.66	≈	130.46	≪	146.7		
BOS11–27	64.41	≫	54.54	<	70.16		54.54	>	49.15	≪	57.95		
BOS11–30	42.87	≈	42.83	≪	56.32		42.83	≫	15.97	≪	30.27		
BOS11–31	88.38	≈	89.66	≈	94.45		89.66	≈	86.84	≈	85.93		
BK12–28	585.42	≈	570.56	≈	570.87		570.56	≈	567.86	≈	566		
D05–18	154.75	≫	138.83	≪	151.81		138.83	≈	136.21	<	141.56		
D05–19	248.85	≈	235.87	≈	242.34		235.87	≈	220.17	<	231.25		
DF11–6	339.47	≈	347.55	≈	316.5		347.55	≈	333.68	≈	341.04		
DF11–7	526.77	≈	523.18	>	492.09		523.18	≈	512.99	≈	525.32		
DF11–8	563.75	≈	552.71	≈	523.57		552.71	≈	551.81	≈	549.33		
DF11–22	707.53	≈	697.75	≈	694.38		697.75	≈	684.04	≈	689.12		
DF11–23	619.29	<	667.99	≈	656.5		667.99	≈	656.29	≈	643.12		
DF11–24	740.3	≈	758.9	≈	720.58		758.9	≈	779.03	≈	786.01		
DF15–4	175.94	≈	180.54	>	168.75		180.54	≫	149.71	≪	174.17		
DF15–5	737.26	≈	726.27	>	674.63		726.27	≈	716.51	≈	689.89		
DF15–20	960.95	>	913.96	≈	900.5		913.96	≈	914.2	≈	931.26		
DF15–21	1234.63	≈	1210.32	≈	1218.06		1210.32	≈	1196.29	≈	1181.18		
DF15–33	1800.02	≫	1671.27	≈	1691.77		1671.27	≈	1714.32	≈	1735.23		
DF15–35	310.13	≈	293.31	≈	295.78		293.31	≈	283.21	≈	285.12		
DRFN08–10	208.83	≈	197.36	≈	213.97		197.36	>	180.96	≈	190.57		
DRFN08–11	264.48	≈	263.16	≈	258.39		263.16	≈	254.11	≈	256.45		
DO09–32	1047.59	≈	1053.04	≈	1049.79		1053.04	≈	1090.22	≈	1086.17		
FY17–25	188.5	≈	183.48	≈	175.59		183.48	≫	161.45	≈	162.3		
FRD12–29	319.45	≈	308.6	≈	320.55		308.6	≈	291.43	≈	295.84		
KS13–12	1761.98	≈	1780.97	>	1694.94		1780.97	≈	1782.82	>	1706.03		
STS13–13	865.67	≈	907.14	>	858.65		907.14	≈	912.8	≈	899.24		
Aoyagi and Frechette (2009)	645.31	≈	646.53	≈	649.53		646.53	≈	694.72	≈	739.24		
Blonski et al. (2011)	713.8	≫	670.62	≪	875.74		670.62	≫	549.45	≪	621.23		
Bruttel and Kamecke (2012)	585.42	≈	570.56	≈	570.87		570.56	≈	567.86	≈	566		
Dal Bó (2005)	407.86	≫	378.95	<	404.78		378.95	>	358.51	<	374.94		
Dal Bó and Fréchette (2011)	3536.73	≈	3589.36	≈	3524.77		3589.36	≈	3533.99	≈	3550.1		
Dal Bó and Fréchette (2015)	5259.64	≫	5037.82	≈	5057.54		5037.82	≈	4991.74	≈	5014.35		
Dreber et al. (2008)	478.09	≈	466.13	≈	482.86		466.13	≫	437.17	≈	449.12		
Duffy and Ochs (2009)	1047.59	≈	1053.04	≈	1049.79		1053.04	≈	1090.22	≈	1086.17		
Fréchette and Yuksel (2017)	188.5	≈	183.48	≈	175.59		183.48	≫	161.45	≈	162.3		
Fudenberg et al. (2012)	319.45	≈	308.6	≈	320.55		308.6	≈	291.43	≈	295.84		
Kagel and Schley (2013)	1761.98	≈	1780.97	>	1694.94		1780.97	≈	1782.82	>	1706.03		
Sherstyuk et al. (2013)	865.67	≈	907.14	>	858.65		907.14	≈	912.8	≈	899.24		
Pooled	**16077.95**	>	**15853.82**	≪	**16335.33**		**15853.82**	≫	**15481.59**	≈	**15573.98**		

(b) Second halves per session

	Best mixture of generalized strategies										Semi-Grim		AD + SG2
	No Switching		Random Switching		Markov Switching		Best Switching						
Specification													
# Models evaluated	4		4		4						1		1
# Pars estimated (by treatment)	32		32		98						3		3
# Parameters accounted for	6–10		6–10		15-35						3		3
AF09–34	998.99	≈	1011.31	≈	989.06		989.06	≈	1085.04	<	1178.07		
BOS11–9	56.83	<	67.33	≈	63.72		63.72	>	50.83	≈	51.64		
BOS11–14	75.9	≈	70.2	<	77.47		77.47	>	61.74	≪	69.39		
BOS11–15	43.38	≫	34.25	≪	55.25		55.25	≫	21.88	≪	37.94		
BOS11–16	170.31	≈	187.28	≈	176.88		176.88	≈	178.84	≈	174.16		
BOS11–17	191.55	≈	192.15	≈	194.05		194.05	≈	183.03	≈	181.55		
BOS11–26	328.04	≈	330.67	≈	340.16		340.16	≈	317.33	≈	317.2		
BOS11–27	214.12	≈	222.33	≈	218.59		218.59	≈	231.72	≈	215.78		
BOS11–30	116.18	≫	109.21	≈	116.12		116.12	≫	93.22	≈	99.81		
BOS11–31	183.27	≈	185.32	≈	188.11		188.11	≈	200.87	≈	188.19		
BK12–28	915.48	≈	926.52	>	879.03		879.03	≈	916.81	≈	903.59		
D05–18	360.73	≈	361.89	≈	355.8		355.8	≈	351.89	≈	353.01		
D05–19	442.98	≈	435.7	≈	431.03		431.03	≈	418.28	≈	423.88		
DF11–6	584.49	≈	615.02	≈	531.24		531.24	≈	599.29	≈	579.61		
DF11–7	1135.74	≈	1179.37	≫	996.23		996.23	≪	1199.27	≈	1168.14		
DF11–8	971.84	≈	971.32	>	874.03		874.03	≪	985.67	≈	956.71		
DF11–22	1348.86	≈	1333.85	>	1269.49		1269.49	≈	1331.85	≈	1303.37		
DF11–23	1105.85	≈	1135.28	>	1043.89		1043.89	≈	1113.94	≈	1113.42		
DF11–24	1076.38	≈	1105.99	≈	1038.16		1038.16	≈	1122.26	≈	1162.51		
DF15–4	298.69	≈	283.7	≈	279.23		279.23	≈	260.02	<	283.22		
DF15–5	1343.25	≈	1361.6	≫	1240.17		1240.17	<	1331.49	≈	1291.93		
DF15–20	1696.3	≈	1689.33	≫	1589.04		1589.04	<	1716.02	≈	1702.32		
DF15–21	2152.39	≈	2212.29	≫	2054.9		2054.9	<	2213.88	≈	2176.02		
DF15–33	3731.12	>	3591.79	≈	3531.59		3531.59	<	3678.61	≈	3654.5		
DF15–35	910.07	>	854.66	≈	870.49		870.49	≈	856.78	≈	842.09		
DRFN08–10	310.63	≈	308.06	≈	311.09		311.09	>	270.35	<	297.4		
DRFN08–11	419.57	≈	427.47	>	404.08		404.08	≈	425.89	≈	423.94		
DO09–32	2401.04	≈	2451.43	≫	2321.8		2321.8	<	2526.02	≈	2569.75		
FY17–25	475.52	≈	454.57	≈	459.64		459.64	≈	473.84	≈	477.72		
FRD12–29	636.43	<	685.17	>	644.13		644.13	≈	668.18	≈	652.38		
KS13–12	2935.17	≈	3003.32	≫	2783.52		2783.52	<	3009.28	>	2831.74		
STS13–13	1358.57	<	1493.91	≫	1325.94		1325.94	<	1507.33	≈	1459.95		
Aoyagi and Frechette (2009)	998.99	≈	1011.31	≈	989.06		989.06	≈	1085.04	<	1178.07		
Blonski et al. (2011)	1450.51	≈	1471.96	≪	1608.54		1608.54	≫	1369.51	≈	1365.7		
Bruttel and Kamecke (2012)	915.48	≈	926.52	>	879.03		879.03	≈	916.81	≈	903.59		
Dal Bó (2005)	807.96	≈	803.27	≈	797.46		797.46	≈	772.3	≈	779.01		
Dal Bó and Fréchette (2011)	6260.83	≈	6383.92	≫	5874.18		5874.18	≪	6368.43	≈	6299.92		
Dal Bó and Fréchette (2015)	10177.65	≈	10037.43	≫	9693.72		9693.72	≪	10074.29	≈	9967.59		
Dreber et al. (2008)	734.98	≈	740.32	≈	728.28		728.28	≈	698.35	≈	723.44		
Duffy and Ochs (2009)	2401.04	≈	2451.43	≫	2321.8		2321.8	<	2526.02	≈	2569.75		
Fréchette and Yuksel (2017)	475.52	≈	454.57	≈	459.64		459.64	≈	473.84	≈	477.72		
Fudenberg et al. (2012)	636.43	<	685.17	>	644.13		644.13	≈	668.18	≈	652.38		
Kagel and Schley (2013)	2935.17	≈	3003.32	≫	2783.52		2783.52	<	3009.28	>	2831.74		
Sherstyuk et al. (2013)	1358.57	<	1493.91	≫	1325.94		1325.94	<	1507.33	≈	1459.95		
Pooled	**29421.33**	≈	**29743.95**	≫	**28851.02**		**28851.02**	≪	**29578.81**	>	**29318.3**		

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 7.

Table 17: 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		
	No Switching		Random Switching		Markov Switching		Best Switching		Semi-Grim		SG + Noise		Best Switching By Treatment		
Specification															
# Models evaluated	8^{32}		8^{32}		8^{32}		3×8^{32}		1		1		$24^{32} \approx 10^{44}$		
# Pars estimated (by treatment)	48		48		180		276		3		4		276		
# Parameters accounted for	3–10		3–10		12-35		3–10		3		4		3–10		
First halves per session															
Aoyagi and Frechette (2009)	645.31	$\approx$	646.53	$\approx$	649.53		646.53	$\approx$	694.72	$>$	646.73	$\approx$	645.31		
Blonski et al. (2011)	677.42	$\gg$	619.43	$\ll$	821.29		619.43	$\gg$	549.45	$\approx$	558.66	$<$	614.47		
Bruttel and Kamecke (2012)	585.42	$\approx$	570.56	$\approx$	570.87		570.56	$\approx$	567.86	$\approx$	559.05	$\approx$	570.56		
Dal Bó (2005)	394.44	$\gg$	365.35	$\ll$	393		365.35	$\approx$	358.51	$\approx$	361.22	$\approx$	365.35		
Dal Bó and Fréchette (2011)	3536.73	$\approx$	3576.34	$\approx$	3524.77		3576.34	$\approx$	3533.99	$>$	3416.54	$\approx$	3471.98		
Dal Bó and Fréchette (2015)	5250.62	$\gg$	5006.33	$\approx$	5049.55		5006.33	$\approx$	4991.74	$\approx$	4935.93	$\approx$	4965.57		
Dreber et al. (2008)	459.11	$\approx$	463.01	$\approx$	477.02		463.01	$>$	437.17	$\approx$	435.65	$\approx$	461.11		
Duffy and Ochs (2009)	1047.59	$\approx$	1053.04	$\approx$	1049.79		1053.04	$\approx$	1090.22	$>$	1017.89	$\approx$	1047.59		
Fréchette and Yuksel (2017)	181.98	$\gg$	161.75	$<$	175.59		161.75	$\approx$	161.45	$\approx$	164.51	$\approx$	161.75		
Fudenberg et al. (2012)	319.45	$\approx$	308.6	$\approx$	320.55		308.6	$\approx$	291.43	$\approx$	288.89	$\approx$	308.6		
Kagel and Schley (2013)	1761.98	$\approx$	1780.97	$>$	1694.94		1780.97	$\approx$	1782.82	$\approx$	1726.06	$\approx$	1694.94		
Sherstyuk et al. (2013)	865.67	$\approx$	907.14	$>$	858.65		907.14	$\approx$	912.8	$\approx$	868.89	$\approx$	858.65		
Pooled	**15951.95**	$\gg$	**15675.49**	$\ll$	**16214.1**		**15675.49**	$>$	**15481.59**	$\gg$	**15125.92**	$\ll$	**15535.6**		
Second halves per session															
Aoyagi and Frechette (2009)	363.58	$\approx$	368.23	$\approx$	368.89		363.58	$\approx$	389.24	$>$	353.94	$\approx$	363.58		
Blonski et al. (2011)	946.2	$>$	912.16	$\ll$	1107.23		946.2	$>$	867.87	$\approx$	868.31	$\approx$	908.25		
Bruttel and Kamecke (2012)	342.17	$\approx$	358.12	$\approx$	347.08		342.17	$\approx$	347.4	$\approx$	342.6	$\approx$	342.17		
Dal Bó (2005)	461.23	$\approx$	445.57	$<$	469.98		461.23	$>$	424.44	$\approx$	429.06	$\approx$	442.59		
Dal Bó and Fréchette (2011)	2737.11	$<$	2865.45	$>$	2721.88		2737.11	$\approx$	2817.31	$\approx$	2694.4	$\approx$	2700.09		
Dal Bó and Fréchette (2015)	5153.82	$\approx$	5067.01	$\approx$	5106.57		5153.82	$\approx$	5043.81	$>$	4902.82	$\approx$	4998.09		
Dreber et al. (2008)	287.49	$\approx$	281.99	$\approx$	299.36		287.49	$\approx$	264.94	$\approx$	268.86	$\approx$	283.38		
Duffy and Ochs (2009)	1381.01	$\approx$	1416.71	$\approx$	1392.49		1381.01	$\approx$	1403.03	$>$	1308.8	$\approx$	1381.01		
Fréchette and Yuksel (2017)	309.63	$\approx$	304.7	$\approx$	308.78		309.63	$\approx$	313.5	$\approx$	278.74	$<$	304.7		
Fudenberg et al. (2012)	373.44	$\approx$	395.32	$\approx$	376.62		373.44	$\approx$	380.75	$\approx$	358.86	$\approx$	373.44		
Kagel and Schley (2013)	1170.12	$\approx$	1224.37	$>$	1143.67		1170.12	$\approx$	1211.37	$\approx$	1153.04	$\approx$	1143.67		
Sherstyuk et al. (2013)	527.09	$\approx$	590.16	$\approx$	567.63		527.09	$\approx$	586.72	$\approx$	564.5	$\approx$	527.09		
Pooled	**14269.01**	$\approx$	**14448.15**	$<$	**14877.81**		**14269.01**	$\approx$	**14159.8**	$\gg$	**13669.82**	$\ll$	**14108.4**		

Note: This table extends Table 11 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.

Table 18: Table 17 by treatments – Best mixtures of pure or generalized strategies in relation to Semi-Grim

(a) First halves per session

	Best mixture of pure or generalized strategies										Best Mixture	
	No Switching		Random Switching		Markov Switching		Best Switching		Semi-Grim		SG + Noise		Best Switching By Treatment
Specification													
# Models evaluated	8^{32}		8^{32}		8^{32}		3×8^{32}		1		1		$24^{32} \approx 10^{44}$
# Pars estimated (by treatment)	48		48		180		276		3		4		276
# Parameters accounted for	3–10		3–10		12-35		3–10		3		4		3–10
AF09–34	645.31	$\approx$	646.53	$\approx$	649.53		646.53	$\approx$	694.72	$>$	646.73	$\approx$	645.31
BOS11–9	39.68	$\approx$	34.86	$\ll$	47.44		34.86	$\gg$	28.34	$\approx$	28.33	$<$	34.86
BOS11–14	52.96	$\approx$	51.25	$\ll$	69.19		51.25	$>$	43.74	$\approx$	45.24	$<$	51.25
BOS11–15	29.87	$\gg$	22.09	$\ll$	43.24		22.09	$\gg$	14.27	$\approx$	15.77	$\ll$	22.09
BOS11–16	106.92	$\approx$	103.63	$\approx$	112.68		103.63	$\approx$	102.31	$\approx$	93.04	$\approx$	103.63
BOS11–17	48.66	$<$	53.62	$\approx$	57.44		53.62	$\approx$	48.33	$\ll$	54.96	$\approx$	48.66
BOS11–26	162.52	$>$	139.66	$<$	153.41		139.66	$\approx$	130.46	$\approx$	128.89	$<$	139.66
BOS11–27	64.41	$>$	52.93	$\ll$	65.67		52.93	$\approx$	49.15	$\approx$	50.65	$\approx$	52.93
BOS11–30	30.38	$\approx$	30.39	$\ll$	43.76		30.39	$\gg$	15.97	$\approx$	17.47	$\ll$	30.38
BOS11–31	88.38	$>$	79.66	$\ll$	94.45		79.66	$\approx$	86.84	$\approx$	84.25	$\approx$	79.66
BK12–28	585.42	$\approx$	570.56	$\approx$	570.87		570.56	$\approx$	567.86	$\approx$	559.05	$\approx$	570.56
D05–18	153.77	$\gg$	136.89	$\ll$	151.01		136.89	$\approx$	136.21	$\approx$	139.6	$\approx$	136.89
D05–19	238.55	$\approx$	226.34	$\approx$	233.48		226.34	$\approx$	220.17	$\approx$	218.78	$\approx$	226.34
DF11–6	339.47	$\approx$	347.55	$\approx$	316.5		347.55	$\approx$	333.68	$\approx$	318.37	$\approx$	316.5
DF11–7	526.77	$\approx$	513.5	$\approx$	492.09		513.5	$\approx$	512.99	$\approx$	501.16	$\approx$	492.09
DF11–8	563.75	$\approx$	552.71	$\approx$	523.57		552.71	$\approx$	551.81	$\approx$	547.22	$\approx$	523.57
DF11–22	707.53	$\approx$	697.75	$\approx$	694.38		697.75	$\approx$	684.04	$\approx$	655.17	$<$	694.38
DF11–23	619.29	$<$	667.99	$\approx$	656.5		667.99	$\approx$	656.29	$\approx$	639.07	$\approx$	619.29
DF11–24	740.3	$\approx$	758.9	$\approx$	720.58		758.9	$\approx$	779.03	$\approx$	734	$\approx$	720.58
DF15–4	170.82	$\approx$	158.62	$\approx$	164.65		158.62	$\approx$	149.71	$\approx$	151.68	$\approx$	158.62
DF15–5	737.26	$\approx$	724.76	$>$	674.63		724.76	$\approx$	716.51	$\approx$	705.07	$\approx$	674.63
DF15–20	960.95	$>$	913.96	$\approx$	900.5		913.96	$\approx$	914.2	$\approx$	901.08	$\approx$	900.5
DF15–21	1234.63	$\approx$	1210.32	$\approx$	1218.06		1210.32	$\approx$	1196.29	$>$	1174.44	$\approx$	1210.32
DF15–33	1800.02	$\gg$	1671.27	$\approx$	1691.77		1671.27	$\approx$	1714.32	$\approx$	1703.43	$\approx$	1671.27
DF15–35	310.13	$\approx$	293.31	$\approx$	295.78		293.31	$\approx$	283.21	$\approx$	276.9	$\approx$	293.31
DRFN08–10	197.04	$\approx$	195.4	$\approx$	208.99		195.4	$\approx$	180.96	$\approx$	182.62	$\approx$	195.4
DRFN08–11	259.69	$\approx$	263.16	$\approx$	258.39		263.16	$\approx$	254.11	$\approx$	250.23	$\approx$	258.39
DO09–32	1047.59	$\approx$	1053.04	$\approx$	1049.79		1053.04	$\approx$	1090.22	$>$	1017.89	$\approx$	1047.59
FY17–25	181.98	$\gg$	161.75	$<$	175.59		161.75	$\approx$	161.45	$\approx$	164.51	$\approx$	161.75
FRD12–29	319.45	$\approx$	308.6	$\approx$	320.55		308.6	$\approx$	291.43	$\approx$	288.89	$\approx$	308.6
KS13–12	1761.98	$\approx$	1780.97	$>$	1694.94		1780.97	$\approx$	1782.82	$\approx$	1726.06	$\approx$	1694.94
STS13–13	865.67	$\approx$	907.14	$>$	858.65		907.14	$\approx$	912.8	$\approx$	868.89	$\approx$	858.65
Aoyagi and Frechette (2009)	645.31	$\approx$	646.53	$\approx$	649.53		646.53	$\approx$	694.72	$>$	646.73	$\approx$	645.31
Blonski et al. (2011)	677.42	$\gg$	619.43	$\ll$	821.29		619.43	$\gg$	549.45	$\approx$	558.66	$<$	614.47
Bruttel and Kamecke (2012)	585.42	$\approx$	570.56	$\approx$	570.87		570.56	$\approx$	567.86	$\approx$	559.05	$\approx$	570.56
Dal Bó (2005)	394.44	$\gg$	365.35	$\ll$	393		365.35	$\approx$	358.51	$\approx$	361.22	$\approx$	365.35
Dal Bó and Fréchette (2011)	3536.73	$\approx$	3576.34	$\approx$	3524.77		3576.34	$\approx$	3533.99	$>$	3416.54	$\approx$	3471.98
Dal Bó and Fréchette (2015)	5250.62	$\gg$	5006.33	$\approx$	5049.55		5006.33	$\approx$	4991.74	$\approx$	4935.93	$\approx$	4965.57
Dreber et al. (2008)	459.11	$\approx$	463.01	$\approx$	477.02		463.01	$>$	437.17	$\approx$	435.65	$\approx$	461.11
Duffy and Ochs (2009)	1047.59	$\approx$	1053.04	$\approx$	1049.79		1053.04	$\approx$	1090.22	$>$	1017.89	$\approx$	1047.59
Fréchette and Yuksel (2017)	181.98	$\gg$	161.75	$<$	175.59		161.75	$\approx$	161.45	$\approx$	164.51	$\approx$	161.75
Fudenberg et al. (2012)	319.45	$\approx$	308.6	$\approx$	320.55		308.6	$\approx$	291.43	$\approx$	288.89	$\approx$	308.6
Kagel and Schley (2013)	1761.98	$\approx$	1780.97	$>$	1694.94		1780.97	$\approx$	1782.82	$\approx$	1726.06	$\approx$	1694.94
Sherstyuk et al. (2013)	865.67	$\approx$	907.14	$>$	858.65		907.14	$\approx$	912.8	$\approx$	868.89	$\approx$	858.65
Pooled	**15951.95**	$\gg$	**15675.49**	$\ll$	**16214.1**		**15675.49**	$>$	**15481.59**	$\gg$	**15125.92**	$\ll$	**15535.6**

(b) Second halves per session

	Best mixture of pure or generalized strategies										Best Mixture	
	No Switching		Random Switching		Markov Switching		Best Switching		Semi-Grim		SG + Noise		Best Switching By Treatment
Specification													
# Models evaluated	8^{32}		8^{32}		8^{32}		3×8^{32}		1		1		$24^{32} \approx 10^{44}$
# Pars estimated (by treatment)	48		48		180		276		3		4		276
# Parameters accounted for	3–10		3–10		12-35		3–10		3		4		3–10
AF09–34	998.99	$\approx$	1011.31	$\approx$	989.06		989.06	$\approx$	1085.04	$>$	1002.44	$\approx$	989.06
BOS11–9	52.34	$\approx$	53.6	$<$	61.92		61.92	$>$	50.83	$\approx$	52.33	$\approx$	52.34
BOS11–14	75.9	$\approx$	70.2	$<$	77.47		77.47	$>$	61.74	$\approx$	63.23	$<$	70.2
BOS11–15	39.65	$\gg$	34.25	$\ll$	53.06		53.06	$\gg$	21.88	$\approx$	24.13	$<$	34.25
BOS11–16	170.31	$\approx$	184.91	$\approx$	176.88		176.88	$\approx$	178.84	$\approx$	159.07	$\approx$	170.31
BOS11–17	191.55	$\approx$	190.05	$\approx$	189.58		189.58	$\approx$	183.03	$\approx$	184.58	$\approx$	189.58
BOS11–26	328.04	$\approx$	330.67	$\approx$	340.16		340.16	$\approx$	317.33	$\approx$	315.97	$\approx$	328.04
BOS11–27	213.5	$\approx$	221.89	$\approx$	218.59		218.59	$\approx$	231.72	$\approx$	229.23	$\approx$	213.5
BOS11–30	110.41	$\gg$	100.79	$\approx$	112.27		112.27	$\gg$	93.22	$\approx$	94.72	$\approx$	100.79
BOS11–31	182.25	$\approx$	185.32	$\approx$	188.11		188.11	$\approx$	200.87	$\approx$	188.4	$\approx$	182.25
BK12–28	915.48	$\approx$	926.52	$>$	879.03		879.03	$\approx$	916.81	$\approx$	902.34	$\approx$	879.03
D05–18	360.73	$\approx$	361.89	$\approx$	355.8		355.8	$\approx$	351.89	$\approx$	352.95	$\approx$	355.8
D05–19	440.24	$\approx$	433.06	$\approx$	431.03		431.03	$\approx$	418.28	$\approx$	409	$\approx$	431.03
DF11–6	584.49	$\approx$	615.02	$\approx$	531.24		531.24	$\approx$	599.29	$\approx$	543.48	$\approx$	531.24
DF11–7	1135.74	$\approx$	1179.37	$\gg$	996.23		996.23	$\ll$	1199.27	$>$	1089.22	$>$	996.23
DF11–8	971.84	$\approx$	971.32	$>$	874.03		874.03	$\ll$	985.67	$\approx$	985.55	$>$	874.03
DF11–22	1348.86	$\approx$	1333.85	$>$	1269.49		1269.49	$\approx$	1331.85	$\approx$	1333.74	$\approx$	1269.49
DF11–23	1105.85	$\approx$	1135.28	$>$	1043.89		1043.89	$\approx$	1113.94	$\approx$	1098.69	$\approx$	1043.89
DF11–24	1076.38	$\approx$	1105.99	$\approx$	1038.16		1038.16	$\approx$	1122.26	$\approx$	1088.71	$\approx$	1038.16
DF15–4	298.69	$\gg$	270.13	$\approx$	272.31		272.31	$\approx$	260.02	$\approx$	262.05	$\approx$	270.13
DF15–5	1343.25	$\approx$	1349.58	$\gg$	1240.17		1240.17	$<$	1331.49	$\approx$	1325.96	$>$	1240.17
DF15–20	1696.3	$\approx$	1689.33	$\gg$	1589.04		1589.04	$<$	1716.02	$\approx$	1677.48	$>$	1589.04
DF15–21	2152.39	$\approx$	2212.29	$\gg$	2054.9		2054.9	$<$	2213.88	$>$	2121.22	$>$	2054.9
DF15–33	3731.12	$>$	3591.79	$\approx$	3531.59		3531.59	$<$	3678.61	$\approx$	3633.99	$\approx$	3531.59
DF15–35	910.07	$>$	854.66	$\approx$	870.49		870.49	$\approx$	856.78	$\approx$	858.57	$\approx$	854.66
DRFN08–10	308.43	$\approx$	306.9	$\approx$	311.09		311.09	$>$	270.35	$\approx$	272.02	$\ll$	306.9
DRFN08–11	419.57	$\approx$	427.47	$>$	404.08		404.08	$\approx$	425.89	$\approx$	427.44	$\approx$	404.08
DO09–32	2401.04	$\approx$	2451.43	$\gg$	2321.8		2321.8	$<$	2526.02	$>$	2344.75	$\approx$	2321.8
FY17–25	475.52	$\approx$	454.57	$\approx$	459.64		459.64	$\approx$	473.84	$\approx$	451.38	$\approx$	454.57
FRD12–29	636.43	$<$	685.17	$>$	644.13		644.13	$\approx$	668.18	$\approx$	632.09	$\approx$	636.43
KS13–12	2935.17	$\approx$	3003.32	$\gg$	2783.52		2783.52	$<$	3009.28	$\approx$	2904.04	$>$	2783.52
STS13–13	1358.57	$<$	1493.91	$\gg$	1325.94		1325.94	$<$	1507.33	$\approx$	1446.27	$>$	1325.94
Aoyagi and Frechette (2009)	998.99	$\approx$	1011.31	$\approx$	989.06		989.06	$\approx$	1085.04	$>$	1002.44	$\approx$	989.06
Blonski et al. (2011)	1413	$\approx$	1425.33	$\ll$	1572.06		1572.06	$\gg$	1369.51	$\approx$	1351.72	$\approx$	1402.96
Bruttel and Kamecke (2012)	915.48	$\approx$	926.52	$>$	879.03		879.03	$\approx$	916.81	$\approx$	902.34	$\approx$	879.03
Dal Bó (2005)	804.43	$\approx$	799.3	$\approx$	797.46		797.46	$\approx$	772.3	$\approx$	764.79	$<$	797.46
Dal Bó and Fréchette (2011)	6260.83	$\approx$	6383.92	$\gg$	5874.18		5874.18	$\ll$	6368.43	$>$	6160.93	$>$	5874.18
Dal Bó and Fréchette (2015)	10177.65	$>$	10006.1	$\gg$	9682.9		9682.9	$\ll$	10074.29	$>$	9902.6	$\gg$	9631.07
Dreber et al. (2008)	731.63	$\approx$	737.99	$\approx$	728.28		728.28	$\approx$	698.35	$\approx$	702.26	$\approx$	718.29
Duffy and Ochs (2009)	2401.04	$\approx$	2451.43	$\gg$	2321.8		2321.8	$<$	2526.02	$>$	2344.75	$\approx$	2321.8
Fréchette and Yuksel (2017)	475.52	$\approx$	454.57	$\approx$	459.64		459.64	$\approx$	473.84	$\approx$	451.38	$\approx$	454.57
Fudenberg et al. (2012)	636.43	$<$	685.17	$>$	644.13		644.13	$\approx$	668.18	$\approx$	632.09	$\approx$	636.43
Kagel and Schley (2013)	2935.17	$\approx$	3003.32	$\gg$	2783.52		2783.52	$<$	3009.28	$\approx$	2904.04	$>$	2783.52
Sherstyuk et al. (2013)	1358.57	$<$	1493.91	$\gg$	1325.94		1325.94	$<$	1507.33	$\approx$	1446.27	$>$	1325.94
Pooled	**29345.07**	$\approx$	**29624.32**	$\gg$	**28779.62**		**28779.62**	$\ll$	**29578.81**	$\gg$	**28711.49**	$>$	**28340.25**

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 7.

Table 19: Is there a single "semi grim" type? Mixture models involving SG

	SG + TFT		SG + AD		3× SG		2× SG + Noise		SG + SG		SG + Noise		Semi-Grim
Specification													
# Models evaluated	1		1		1		1		1		1		1
# Pars estimated (by treatment)	5		5		11		8		7		4		3
# Parameters accounted for	5		5		11		8		7		4		3
First halves per session													
Aoyagi and Frechette (2009)	698.37	$\approx$	698.36	$>$	650.89	$\approx$	647.47	$\approx$	650.83	$\approx$	646.73	$\approx$	694.72
Blonski et al. (2011)	614.21	$\ll$	663.28	$\ll$	743.15	$\gg$	667.95	$\gg$	642.5	$\gg$	558.66	$\approx$	549.45
Bruttel and Kamecke (2012)	571.1	$\approx$	567.84	$\approx$	556.46	$\approx$	554.86	$\approx$	553.59	$\approx$	559.05	$\approx$	567.86
Dal Bó (2005)	386.98	$\approx$	378.87	$\approx$	401.23	$\gg$	385.12	$\approx$	379.63	$>$	361.22	$\approx$	358.51
Dal Bó and Fréchette (2011)	3538.27	$\approx$	3488.81	$>$	3327.75	$\approx$	3325.99	$\approx$	3335.4	$<$	3416.54	$<$	3533.99
Dal Bó and Fréchette (2015)	5047.74	$\approx$	5024.63	$\approx$	5004.06	$>$	4946.44	$\approx$	4958.73	$\approx$	4935.93	$\approx$	4991.74
Dreber et al. (2008)	443.9	$\approx$	442.47	$\approx$	454.03	$\approx$	442.03	$\approx$	438.35	$\approx$	435.65	$\approx$	437.17
Duffy and Ochs (2009)	1099.57	$\approx$	1084.74	$>$	1019.33	$\approx$	1018.29	$\approx$	1016.91	$\approx$	1017.89	$<$	1090.22
Fréchette and Yuksel (2017)	180.54	$\gg$	165.94	$\ll$	187.62	$\approx$	183.63	$\approx$	179.6	$>$	164.51	$\approx$	161.45
Fudenberg et al. (2012)	297.99	$\approx$	292.47	$\approx$	290.04	$\approx$	289.31	$\approx$	287.31	$\approx$	288.89	$\approx$	291.43
Kagel and Schley (2013)	1777.58	$>$	1689.17	$\approx$	1622.36	$\approx$	1680.43	$\approx$	1677.99	$\approx$	1726.06	$\approx$	1782.82
Sherstyuk et al. (2013)	904.29	$\approx$	887.79	$\approx$	868.66	$\approx$	870.16	$\approx$	868.11	$\approx$	868.89	$\approx$	912.8
Pooled	**15742.91**	$>$	**15566.74**	$\approx$	**15526.79**	$\gg$	**15303.48**	$\approx$	**15244.28**	$\approx$	**15125.92**	$\ll$	**15481.59**
Second halves per session													
Aoyagi and Frechette (2009)	397.94	$\approx$	392.87	$\approx$	363.39	$>$	360.1	$\approx$	358.14	$>$	353.94	$\approx$	389.24
Blonski et al. (2011)	940.75	$\approx$	914.98	$<$	985.46	$\gg$	927.39	$\gg$	906.87	$\approx$	868.31	$\approx$	868.72
Bruttel and Kamecke (2012)	337.22	$\approx$	349.32	$\gg$	312.72	$\approx$	316.23	$\approx$	322.87	$\approx$	342.6	$\approx$	347.4
Dal Bó (2005)	444.73	$\approx$	437.71	$\ll$	476.01	$>$	458.8	$\approx$	455.9	$\gg$	429.06	$\approx$	424.44
Dal Bó and Fréchette (2011)	2798.46	$>$	2640.88	$\approx$	2554.16	$\approx$	2556.55	$\approx$	2545.23	$\ll$	2694.4	$<$	2821.38
Dal Bó and Fréchette (2015)	5053.74	$\approx$	5032.59	$\approx$	4946.96	$\approx$	4909.92	$\approx$	4879.57	$\approx$	4902.82	$<$	5043.81
Dreber et al. (2008)	262.93	$\approx$	275.84	$\approx$	283.45	$>$	272.19	$\approx$	268.1	$\approx$	268.86	$\approx$	265.11
Duffy and Ochs (2009)	1414.84	$\approx$	1399.65	$>$	1288.79	$\approx$	1284.98	$\approx$	1284.59	$\approx$	1308.8	$<$	1403.03
Fréchette and Yuksel (2017)	319.44	$\approx$	319.78	$\approx$	283.19	$\approx$	278.98	$\approx$	277.03	$\approx$	278.74	$<$	313.5
Fudenberg et al. (2012)	388.9	$>$	366.79	$\approx$	345.11	$\approx$	345.11	$\approx$	348.86	$\approx$	358.86	$\approx$	380.75
Kagel and Schley (2013)	1200.87	$>$	1126.89	$\approx$	1065.1	$\approx$	1070.99	$\approx$	1111.19	$\approx$	1153.04	$\approx$	1211.37
Sherstyuk et al. (2013)	567.48	$\approx$	550.64	$\approx$	518	$\approx$	518.83	$\approx$	544.85	$\approx$	564.5	$\approx$	586.72
Pooled	**14309.69**	$\gg$	**13990.32**	$\approx$	**13823.56**	$\gg$	**13591.86**	$\approx$	**13558.53**	$\approx$	**13669.82**	$\ll$	**14164.9**

Note: This table verifies a number of possible mixtures involving Semi-Grim types as a robustness check for the sufficiency of focusing 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 20: Table 19 by treatments – Is there a single "semi grim" type? Mixture models involving SG

(a) First halves per session

	SG + TFT		SG + AD		3× SG		2× SG + Noise		SG + SG		SG + Noise		Semi-Grim
Specification													
# Models evaluated	1		1		1		1		1		1		1
# Pars estimated (by treatment)	5		5		11		8		7		4		3
# Parameters accounted for	5		5		11		8		7		4		3
AF09–34	698.37	≈	698.36	>	650.89	≈	647.47	≈	650.83	≈	646.73	≈	694.72
BOS11–9	37.72	≈	36.71	≈	36.42	≫	31.91	≈	30.41	≈	28.33	≈	28.34
BOS11–14	46.74	≪	59.53	≈	62.1	>	57.42	≈	55.89	≫	45.24	≈	43.74
BOS11–15	17.7	≪	30.49	≪	34.31	≫	29.77	>	28.27	≫	15.77	>	14.27
BOS11–16	105.34	≈	108.05	≈	102.21	>	97.29	≈	95.79	>	93.04	≈	102.31
BOS11–17	51.6	≈	54.39	<	62.12	≈	58.67	>	55.43	≈	54.96	>	48.33
BOS11–26	138.93	<	145.92	≈	143.32	>	135.62	≈	133.77	≈	128.89	≈	130.46
BOS11–27	57.08	≈	57.83	≪	66.2	≫	61.51	≈	60.21	>	50.65	≈	49.15
BOS11–30	19.19	≪	31.51	≪	42.82	≫	36.78	>	35.23	≫	17.47	≫	15.97
BOS11–31	89.84	≈	88.76	≈	83.49	≈	78.88	≈	77.39	≈	84.25	≈	86.84
BK12–28	571.1	≈	567.84	≈	556.46	≈	554.86	≈	553.59	≈	559.05	≈	567.86
D05–18	152.64	≫	140.51	≈	147.72	≫	137.32	≈	135.38	≈	139.6	≈	136.21
D05–19	230.8	≈	234.82	≈	245.71	≈	242.13	≈	239.28	>	218.78	≈	220.17
DF11–6	340.51	≈	342.93	≈	318	≈	315.22	≈	312.8	≈	318.37	≈	333.68
DF11–7	514.24	≈	510.31	≈	497	≈	493.05	≈	489.18	≈	501.16	≈	512.99
DF11–8	555.59	≈	521.04	≈	510.54	≈	513.57	≈	511.26	<	547.22	≈	551.81
DF11–22	683.89	≈	690.37	>	638.31	≈	641.64	≈	649.47	≈	655.17	≈	684.04
DF11–23	652.77	≈	625.94	≈	589.27	≈	588.49	≈	605.97	≈	639.07	≈	656.29
DF11–24	764.34	≈	771.29	≈	715.4	≈	730.93	≈	729.03	≈	734	≈	779.03
DF15–4	160.08	<	174.81	≪	190.8	>	179.18	≈	177.25	≫	151.68	≈	149.71
DF15–5	729.79	≈	693.23	≈	694.48	≈	688.23	≈	684.87	≈	705.07	≈	716.51
DF15–20	915.67	≈	936	≈	903.1	≈	898.57	≈	895.31	≈	901.08	≈	914.2
DF15–21	1214.46	≫	1167.96	≈	1149.36	≈	1139.86	≈	1173.16	≈	1174.44	≈	1196.29
DF15–33	1713.96	≈	1740.36	≈	1712.55	≈	1708.28	≈	1705.28	≈	1703.43	≈	1714.32
DF15–35	284.62	≈	283.12	≈	289.62	≈	285.66	≈	282.04	≈	276.9	≈	283.21
DRFN08–10	185.08	≈	185.52	≈	190.34	≈	185.08	≈	183.64	≈	182.62	≈	180.96
DRFN08–11	255.32	≈	253.45	≈	255.98	≈	251.35	≈	249.8	≈	250.23	≈	254.11
DO09–32	1099.57	≈	1084.74	>	1019.33	≈	1018.29	≈	1016.91	≈	1017.89	<	1090.22
FY17–25	180.54	≫	165.94	≪	187.62	≈	183.63	≈	179.6	>	164.51	≈	161.45
FRD12–29	297.99	≈	292.47	≈	290.04	≈	289.31	≈	287.31	≈	288.89	≈	291.43
KS13–12	1777.58	>	1689.17	≈	1622.36	≈	1680.43	≈	1677.99	≈	1726.06	≈	1782.82
STS13–13	904.29	≈	887.79	≈	868.66	≈	870.16	≈	868.11	≈	868.89	≈	912.8
Aoyagi and Frechette (2009)	698.37	≈	698.36	>	650.89	≈	647.47	≈	650.83	≈	646.73	≈	694.72
Blonski et al. (2011)	614.21	≪	663.28	≪	743.15	≫	667.95	≫	642.5	≫	558.66	≈	549.45
Bruttel and Kamecke (2012)	571.1	≈	567.84	≈	556.46	≈	554.86	≈	553.59	≈	559.05	≈	567.86
Dal Bó (2005)	386.98	≈	378.87	≈	401.23	≫	385.12	≈	379.63	>	361.22	≈	358.51
Dal Bó and Fréchette (2011)	3538.27	≈	3488.81	>	3327.75	≈	3325.99	≈	3335.4	<	3416.54	<	3533.99
Dal Bó and Fréchette (2015)	5047.74	≈	5024.63	≈	5004.06	>	4946.44	≈	4958.73	≈	4935.93	≈	4991.74
Dreber et al. (2008)	443.9	≈	442.47	≈	454.03	≈	442.03	≈	438.35	≈	435.65	≈	437.17
Duffy and Ochs (2009)	1099.57	≈	1084.74	>	1019.33	≈	1018.29	≈	1016.91	≈	1017.89	<	1090.22
Fréchette and Yuksel (2017)	180.54	≫	165.94	≪	187.62	≈	183.63	≈	179.6	>	164.51	≈	161.45
Fudenberg et al. (2012)	297.99	≈	292.47	≈	290.04	≈	289.31	≈	287.31	≈	288.89	≈	291.43
Kagel and Schley (2013)	1777.58	>	1689.17	≈	1622.36	≈	1680.43	≈	1677.99	≈	1726.06	≈	1782.82
Sherstyuk et al. (2013)	904.29	≈	887.79	≈	868.66	≈	870.16	≈	868.11	≈	868.89	≈	912.8
Pooled	**15742.91**	>	**15566.74**	≈	**15526.79**	≫	**15303.48**	≈	**15244.28**	≈	**15125.92**	≪	**15481.59**

(b) Second halves per session

	SG + TFT		SG + AD		3× SG		2× SG + Noise		SG + SG		SG + Noise		Semi-Grim
Specification													
# Models evaluated	1		1		1		1		1		1		1
# Pars estimated (by treatment)	5		5		11		8		7		4		3
# Parameters accounted for	5		5		11		8		7		4		3
AF09–34	1089.07	≈	1088.68	>	990.69	≈	991.1	≈	1004.14	≈	1002.44	≈	1085.04
BOS11–9	53.89	≈	53.88	<	64.12	>	59.61	≈	58.1	>	52.33	≈	50.83
BOS11–14	64.87	≪	72.35	≪	81.52	≈	76.99	≈	75.39	>	63.23	≈	61.74
BOS11–15	27.62	≪	40.96	>	39.5	≫	35	>	33.53	≫	24.13	>	22.64
BOS11–16	183.6	≈	177.15	≈	166.77	≈	162.08	≈	160.58	≈	159.07	≈	178.84
BOS11–17	180.21	≈	181.35	≈	190.99	≈	186.36	≈	184.86	≈	184.58	≈	183.03
BOS11–26	316.74	≈	303.4	≈	295.8	≈	303.64	≈	301.8	≈	315.97	≈	317.33
BOS11–27	236.3	≈	216.62	≈	206.64	≈	211.44	≈	211.81	≈	229.23	≈	231.72
BOS11–30	96.22	≈	102.16	≈	106.32	>	101.82	>	100.32	>	94.72	>	93.22
BOS11–31	205.18	≈	191.19	≈	173.37	≈	168.88	≈	167.41	≈	188.4	≈	200.87
BK12–28	908.74	≈	904.92	≈	865.47	≈	871.79	≈	884.68	≈	902.34	≈	916.81
D05–18	355.5	≈	357.75	≈	371.46	≈	365.36	≈	362.37	≈	352.95	≈	351.89
D05–19	424.87	≈	426.93	≈	429.94	≈	423.33	≈	420.68	≈	409	≈	418.28
DF11–6	592.73	≈	550.5	≈	505.61	≈	531.1	≈	529.38	≈	543.48	≈	599.29
DF11–7	1175.92	≈	1122.24	≈	1076.72	≈	1033.39	<	1080.6	≈	1089.22	<	1199.27
DF11–8	998.86	≈	925.57	≈	888.74	≈	883.02	≈	881.18	≪	985.55	≈	985.67
DF11–22	1299.41	≈	1280.15	≈	1252.38	≈	1259.53	≈	1257.63	<	1333.74	≈	1331.85
DF11–23	1107.01	≈	1076.91	≈	1018.94	≈	1017.54	≈	1030.94	≈	1098.69	≈	1113.94
DF11–24	1098.92	≈	1126.04	>	1031.22	≈	1031.12	≈	1029.26	≈	1088.71	≈	1122.26
DF15–4	260.03	<	276.63	≈	278.3	≈	271.52	≈	269.36	≈	262.05	≈	260.02
DF15–5	1343.78	≈	1288.52	≈	1263.24	≈	1261.59	≈	1261.3	<	1325.96	≈	1331.49
DF15–20	1685.68	≈	1666.64	≈	1638.48	≈	1641.52	≈	1658.8	≈	1677.48	≈	1716.02
DF15–21	2221.3	>	2133.97	≈	2092.08	≈	2094.42	≈	2097.89	≈	2121.22	≈	2213.88
DF15–33	3602.23	≈	3627.79	>	3508.8	≈	3550.24	≈	3547.85	<	3633.99	≈	3678.61
DF15–35	845.37	≈	823.8	≈	841.54	≈	836.16	≈	834.42	≈	858.57	≈	856.78
DRFN08–10	273.7	≈	274.83	≈	269.72	≈	264.73	≈	263.18	≈	272.02	≈	270.35
DRFN08–11	423.75	≈	425.89	≈	404.27	≈	405.55	≈	404	≈	427.44	≈	425.89
DO09–32	2524.24	≈	2510.62	>	2305.09	≈	2300.39	≈	2311.91	≈	2344.75	<	2526.02
FY17–25	476.73	≈	476.04	≈	444.24	≈	441.83	≈	443.06	≈	451.38	≈	473.84
FRD12–29	657.64	≈	645.3	>	589.46	≈	608.61	≈	613.78	≈	632.09	≈	668.18
KS13–12	2984.62	>	2795.09	>	2642.29	≈	2711.18	≈	2804.08	≈	2904.04	≈	3009.28
STS13–13	1461.59	≈	1429.53	≈	1368.44	≈	1411.21	≈	1430.92	≈	1446.27	≈	1507.33
Aoyagi and Frechette (2009)	1089.07	≈	1088.68	>	990.69	≈	991.1	≈	1004.14	≈	1002.44	≈	1085.04
Blonski et al. (2011)	1414.72	≈	1389.14	≈	1435.21	>	1385.94	>	1363.89	≈	1351.72	≈	1370.26
Bruttel and Kamecke (2012)	908.74	≈	904.92	≈	865.47	≈	871.79	≈	884.68	≈	902.34	≈	916.81
Dal Bó (2005)	783.92	≈	788.23	≈	809.2	≈	794.36	≈	788.01	≈	764.79	≈	772.3
Dal Bó and Fréchette (2011)	6299.78	≈	6108.34	>	5832.86	≈	5798.78	≈	5846.68	≪	6160.93	<	6368.43
Dal Bó and Fréchette (2015)	9987.54	≈	9846.51	≈	9686.58	≈	9702.12	≈	9710.44	<	9902.6	≈	10074.29
Dreber et al. (2008)	700.95	≈	704.23	≈	681.69	≈	675.88	≈	672.08	≈	702.26	≈	698.35
Duffy and Ochs (2009)	2524.24	≈	2510.62	>	2305.09	≈	2300.39	≈	2311.91	≈	2344.75	<	2526.02
Fréchette and Yuksel (2017)	476.73	≈	476.04	≈	444.24	≈	441.83	≈	443.06	≈	451.38	≈	473.84
Fudenberg et al. (2012)	657.64	≈	645.3	>	589.46	≈	608.61	≈	613.78	≈	632.09	≈	668.18
Kagel and Schley (2013)	2984.62	>	2795.09	>	2642.29	≈	2711.18	≈	2804.08	≈	2904.04	≈	3009.28
Sherstyuk et al. (2013)	1461.59	≈	1429.53	≈	1368.44	≈	1411.21	≈	1430.92	≈	1446.27	≈	1507.33
Pooled	**29471.93**	≫	**28869**	≫	**28052.44**	≈	**27984.99**	≈	**28128.99**	≪	**28711.49**	≪	**29579.56**

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 7.

Appendix D: Robustness checks for Section 4

The tables in this section replicate the tables presented in the Section 4, again provide robustness checks and the results treatment-by-treatment.

- Table 21 shows aggregate state-wise cooperation rates for different lagged histories (cooperation or defection of the opponent in $t-2$) *TFT-Scheme*.
- Table 22 shows aggregate state-wise cooperation rates for different lagged histories (joint cooperation or not in $t-2$) *Grim-Scheme*.
- Table 24 compares different models containing semi-grim to models containing pure strategies assuming no-switching behavior.
- Table 26 compares different models containing semi-grim to models containing pure strategies assuming random-switching behavior.
- Table 28 compares different models containing modifications of semi-grim.
- Table 30 compares different models containing prototypical strategies derived from strategies discussed in previous literature in a No-Switching model.
- Table 32 compares different two parameter versions of semi-grim with models containing prototypical strategies. The memory-2 level follows a *Grim-Scheme* if applicable
- Table 33 compares different two parameter versions of semi-grim with models containing prototypical strategies. The memory-2 level follows a *TFT-Scheme* if applicable
- Table 34 examines all mixtures of Semi-Grim with pure or generalized pure strategies as secondary components (robustness check for Table 4).

Table 21: 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 1, with the obvious adaptation that the Holm-Bonferroni correction now applies to all eight tests per data set.

Table 22: 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 1, with the obvious adaptation that the Holm-Bonferroni correction now applies to all eight tests per data set.

Table 23: Table 22 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/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/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 24: 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)	743.26	$\approx$	742.13	$\gg$	694.72	$<$	763.74	$\approx$	791.34	$\approx$	800.02
Blonski et al. (2011)	713.93	$\gg$	585.39	$\gg$	549.45	$\ll$	681.8	$<$	741.88	$\gg$	707.94
Bruttel and Kamecke (2012)	569.53	$\approx$	570.41	$\approx$	567.86	$\approx$	588.53	$\approx$	588.53	$\approx$	603.31
Dal Bó (2005)	418.65	$\gg$	364.64	$\approx$	358.51	$<$	388.99	$\approx$	393.09	$\approx$	394.31
Dal Bó and Fréchette (2011)	3576.79	$\approx$	3594.64	$\approx$	3533.99	$\ll$	3837.37	$\approx$	3836.86	$<$	3934.11
Dal Bó and Fréchette (2015)	5259.51	$\gg$	5006.42	$\approx$	4991.74	$\ll$	5531.56	$\approx$	5552.56	$\approx$	5595.28
Dreber et al. (2008)	455.42	$\approx$	451.44	$\approx$	437.17	$\approx$	462.66	$\approx$	470.47	$\approx$	462.66
Duffy and Ochs (2009)	1109.12	$\approx$	1089.19	$\approx$	1090.22	$\approx$	1100.13	$\approx$	1102.62	$\approx$	1132.22
Fréchette and Yuksel (2017)	169.83	$\approx$	164.74	$\approx$	161.45	$\ll$	181.95	$\approx$	188.02	$\approx$	181.95
Fudenberg et al. (2012)	313.22	$\approx$	298.69	$\approx$	291.43	$\ll$	357.33	$\approx$	366.77	$\approx$	375.41
Kagel and Schley (2013)	1739.6	$\approx$	1787.59	$\approx$	1782.82	$\approx$	1805.94	$\approx$	1805.94	$\approx$	1818.45
Sherstyuk et al. (2013)	918.44	$\approx$	924.56	$\approx$	912.8	$\approx$	941.91	$\approx$	941.91	$\approx$	960.82
Pooled	**16242.61**	$\gg$	**15762.21**	$\gg$	**15481.59**	$\ll$	**16803.32**	$<$	**16962.38**	$\approx$	**17075.91**
Second halves per session											
Aoyagi and Frechette (2009)	437.99	$\approx$	433.04	$\gg$	389.24	$\ll$	476.75	$\approx$	484.39	$\approx$	487.15
Blonski et al. (2011)	988.68	$\gg$	888.46	$\approx$	867.87	$\ll$	990.66	$\approx$	1084.43	$\gg$	1067.21
Bruttel and Kamecke (2012)	333.93	$\approx$	342.01	$\approx$	347.4	$\approx$	316.34	$\approx$	316.34	$\approx$	343.43
Dal Bó (2005)	475.66	$\gg$	422.93	$\approx$	424.44	$<$	461.93	$\approx$	470.29	$\approx$	464.4
Dal Bó and Fréchette (2011)	2758.96	$<$	2842.94	$\approx$	2817.31	$\approx$	2881.22	$\approx$	2886.16	$\ll$	3251.04
Dal Bó and Fréchette (2015)	5109.07	$>$	5027.27	$\approx$	5043.81	$\ll$	5564.25	$\approx$	5586.2	$\ll$	5853.42
Dreber et al. (2008)	271.57	$\approx$	271.55	$\approx$	264.94	$\approx$	287.56	$<$	293.87	$\approx$	287.56
Duffy and Ochs (2009)	1462.19	$\approx$	1446.73	$>$	1403.03	$\ll$	1617.75	$\approx$	1617.75	$\approx$	1661.55
Fréchette and Yuksel (2017)	348.73	$\gg$	315.35	$\approx$	313.5	$<$	356.1	$\approx$	360.69	$\approx$	356.1
Fudenberg et al. (2012)	386.55	$\approx$	389.34	$\approx$	380.75	$<$	445.16	$\approx$	447.17	$\approx$	476.66
Kagel and Schley (2013)	1161.5	$\approx$	1208.7	$\approx$	1211.37	$\approx$	1169.29	$\approx$	1169.29	$<$	1274.75
Sherstyuk et al. (2013)	551.82	$<$	596.62	$\approx$	586.72	$\approx$	583.79	$\approx$	583.79	$\ll$	691.06
Pooled	**14541.97**	$>$	**14367.31**	$\gg$	**14159.8**	$\ll$	**15319.22**	$\approx$	**15482.74**	$\ll$	**16323.75**

Note: Relation signs, bootstrap procedure, and derived p-values are exactly as above, see Table 1. Pure M1 refers to TFT, Grim, and AD. G2 denotes Grim2. For definitions of the strategies see Table 5.

Table 25: Table 24 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	743.26	≈	742.13	≫	694.72	<	763.74	≈	791.34	≈	800.02
BOS11–9	44.14	≫	29.54	≈	28.34	≪	47.65	≈	50.64	>	47.65
BOS11–14	51.96	>	45.65	≈	43.74	≪	61.4	<	64.39	≈	61.4
BOS11–15	33.17	≫	17.23	≫	14.27	≪	29.85	<	32.85	>	29.85
BOS11–16	114.83	≫	102.32	≈	102.31	≈	109.85	≈	112.82	≈	109.85
BOS11–17	71.23	≫	54.59	≈	48.33	≈	48.64	≈	51.05	≈	48.64
BOS11–26	158.29	≫	136.72	≈	130.46	≪	162.83	≈	163.62	≈	162.83
BOS11–27	55.81	≫	46.89	≈	49.15	<	59.82	≈	59.82	≈	67.15
BOS11–30	36.26	≫	19.4	≫	15.97	≪	30.37	<	33.37	≫	30.37
BOS11–31	78.14	≈	82.99	≈	86.84	≈	95.58	≈	123.24	≈	120.15
BK12–28	569.53	≈	570.41	≈	567.86	≈	588.53	≈	588.53	≈	603.31
D05–18	161.84	≫	140.24	≈	136.21	≪	153.73	≈	156.94	≈	153.73
D05–19	251.84	≫	220.85	≈	220.17	≈	232.61	≈	232.61	≈	238.46
DF11–6	346.27	≈	337.83	≈	333.68	≈	368.96	≈	368.96	≈	389.67
DF11–7	530.09	≈	525.23	≈	512.99	<	565.55	≈	565.55	≈	566.74
DF11–8	525.4	<	557.31	≈	551.81	≈	568.73	≈	568.73	<	613.19
DF11–22	693.26	≈	685.45	≈	684.04	<	755.98	≈	755.98	≈	758.43
DF11–23	653.3	≈	664.99	≈	656.29	≈	718.87	≈	718.87	≈	746.73
DF11–24	790.78	≈	796.91	≈	779.03	≈	831.47	≈	831.85	≈	843.2
DF15–4	189.61	≫	160.51	≈	149.71	<	170.74	≈	174.65	≈	170.74
DF15–5	785.26	≫	716.06	≈	716.51	≈	771.72	≈	772.27	≈	775.61
DF15–20	955.58	≫	916.66	≈	914.2	<	999.68	≈	999.68	≈	1005.75
DF15–21	1228.02	>	1201.29	≈	1196.29	≪	1335.2	≈	1335.42	≈	1389.42
DF15–33	1766.92	≫	1695.1	≈	1714.32	≪	1891.17	≈	1907.06	≈	1900.44
DF15–35	293.29	≈	287.64	≈	283.21	≪	334.32	≈	334.32	≈	335.83
DRFN08–10	195.66	≈	191.06	≈	180.96	<	200.91	≈	204.24	≈	200.91
DRFN08–11	254.85	≈	256.88	≈	254.11	≈	259.65	≈	262.72	≈	259.65
DO09–32	1109.12	≈	1089.19	≈	1090.22	≈	1100.13	≈	1102.62	≈	1132.22
FY17–25	169.83	≈	164.74	≈	161.45	≪	181.95	≈	188.02	≈	181.95
FRD12–29	313.22	≈	298.69	≈	291.43	≪	357.33	≈	366.77	≈	375.41
KS13–12	1739.6	≈	1787.59	≈	1782.82	≈	1805.94	≈	1805.94	≈	1818.45
STS13–13	918.44	≈	924.56	≈	912.8	≈	941.91	≈	941.91	≈	960.82
Aoyagi and Frechette (2009)	743.26	≈	742.13	≫	694.72	<	763.74	≈	791.34	≈	800.02
Blonski et al. (2011)	713.93	≫	585.39	≫	549.45	≪	681.8	<	741.88	≫	707.94
Bruttel and Kamecke (2012)	569.53	≈	570.41	≈	567.86	≈	588.53	≈	588.53	≈	603.31
Dal Bó (2005)	418.65	≫	364.64	≈	358.51	<	388.99	≈	393.09	≈	394.31
Dal Bó and Fréchette (2011)	3576.79	≈	3594.64	≈	3533.99	≪	3837.37	≈	3836.86	<	3934.11
Dal Bó and Fréchette (2015)	5259.51	≫	5006.42	≈	4991.74	≪	5531.56	≈	5552.56	≈	5595.28
Dreber et al. (2008)	455.42	≈	451.44	≈	437.17	≈	462.66	≈	470.47	≈	462.66
Duffy and Ochs (2009)	1109.12	≈	1089.19	≈	1090.22	≈	1100.13	≈	1102.62	≈	1132.22
Fréchette and Yuksel (2017)	169.83	≈	164.74	≈	161.45	≪	181.95	≈	188.02	≈	181.95
Fudenberg et al. (2012)	313.22	≈	298.69	≈	291.43	≪	357.33	≈	366.77	≈	375.41
Kagel and Schley (2013)	1739.6	≈	1787.59	≈	1782.82	≈	1805.94	≈	1805.94	≈	1818.45
Sherstyuk et al. (2013)	918.44	≈	924.56	≈	912.8	≈	941.91	≈	941.91	≈	960.82
Pooled	**16242.61**	≫	**15762.21**	≫	**15481.59**	≪	**16803.32**	<	**16962.38**	≈	**17075.91**

(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	1181.22	≈	1181.85	≫	1085.04	<	1226.63	≈	1270.08	≈	1294.67
BOS11–9	62.29	>	51.41	≈	50.83	≈	52.32	≈	55.32	≈	52.32
BOS11–14	76.71	≫	63.05	≈	61.74	≪	82.94	≈	85.94	≈	82.94
BOS11–15	43.05	≫	27.05	≫	21.88	≪	39.63	<	42.63	>	39.63
BOS11–16	185.97	≈	180.95	≈	178.84	≈	179.08	≈	182.08	≈	179.08
BOS11–17	195.91	≈	192.93	≈	183.03	≈	191.79	≈	191.8	≈	191.79
BOS11–26	333.85	≈	319.54	≈	317.33	≈	356.49	≈	360.18	≈	356.49
BOS11–27	212.67	≈	214.41	≈	231.72	≈	227.96	≈	264.18	≈	274.26
BOS11–30	108.28	≫	94.64	≈	93.22	<	110.39	≈	113.38	>	110.39
BOS11–31	165.71	≈	185.95	≈	200.87	≈	200.12	≈	258.45	≈	264.92
BK12–28	891.89	≈	915.39	≈	916.81	≈	912.08	≈	912.08	≈	954.26
D05–18	368.35	>	356.04	≈	351.89	≈	365.9	≈	371.68	≈	365.9
D05–19	431.03	>	409.37	≈	418.28	≈	434.87	≈	434.87	≈	440.2
DF11–6	582.08	≈	590.39	≈	599.29	≈	633.64	≈	633.64	≈	699.58
DF11–7	1151.03	<	1197.17	≈	1199.27	≈	1202.39	≈	1202.39	≈	1242.46
DF11–8	924.45	≪	996.09	≈	985.67	≈	1062.54	≈	1076.28	<	1131.37
DF11–22	1294.96	≈	1327.01	≈	1331.85	<	1448.27	≈	1449.44	≈	1488.67
DF11–23	1099.92	≈	1135.01	≈	1113.94	≈	1185.1	≈	1191.64	≈	1294.05
DF11–24	1163.3	≈	1159.28	≈	1122.26	<	1260.31	≈	1260.31	≈	1296.29
DF15–4	298.7	≫	271.14	≈	260.02	<	299.84	≈	303.76	≈	299.84
DF15–5	1336.56	≈	1323.6	≈	1331.49	<	1424.58	≈	1428.82	≈	1445.66
DF15–20	1696.48	≈	1706.95	≈	1716.02	≈	1813.68	≈	1813.68	≈	1859.13
DF15–21	2186.41	≈	2233.11	≈	2213.88	≪	2409.31	≈	2420.98	<	2536.56
DF15–33	3618.5	≈	3612.75	<	3678.61	≪	4094.8	≈	4104.19	≈	4132.29
DF15–35	845.68	≈	858.27	≈	856.78	≪	1007.17	≈	1007.17	≈	1060.45
DRFN08–10	295.96	≈	289.36	≈	270.35	≪	313.54	≈	316.88	≈	313.54
DRFN08–11	413.33	≈	426.29	≈	425.89	≈	436.15	≈	439.24	≈	436.15
DO09–32	2559.15	≈	2571.83	≈	2526.02	≪	2779.15	≈	2782.83	≈	2823.45
FY17–25	497.71	≈	473.7	≈	473.84	≈	504.26	≈	508.97	≈	504.26
FRD12–29	658.78	≈	680.53	≈	668.18	<	765.61	≈	779.92	≈	826.75
KS13–12	2880.11	<	3005.9	≈	3009.28	≈	3032.91	≈	3032.91	≈	3114.31
STS13–13	1472.88	≈	1528.33	≈	1507.33	≈	1536.36	≈	1536.36	<	1624.89
Aoyagi and Frechette (2009)	1181.22	≈	1181.85	≫	1085.04	<	1226.63	≈	1270.08	≈	1294.67
Blonski et al. (2011)	1454.55	≫	1380.01	≈	1369.51	<	1478.83	≈	1604.02	≫	1581.88
Bruttel and Kamecke (2012)	891.89	≈	915.39	≈	916.81	≈	912.08	≈	912.08	≈	954.26
Dal Bó (2005)	804.34	≫	768.96	≈	772.3	≈	803.42	≈	810.09	≈	808.22
Dal Bó and Fréchette (2011)	6253.44	≪	6431.87	≈	6368.43	≪	6823.8	≈	6840.62	≪	7168.57
Dal Bó and Fréchette (2015)	10023.16	≈	10034.99	≈	10074.29	≪	11078.13	≈	11107.75	<	11351.42
Dreber et al. (2008)	714.19	≈	719.15	≈	698.35	<	751.8	≈	759.62	≈	751.8
Duffy and Ochs (2009)	2559.15	≈	2571.83	≈	2526.02	≪	2779.15	≈	2782.83	≈	2823.45
Fréchette and Yuksel (2017)	497.71	≈	473.7	≈	473.84	≈	504.26	≈	508.97	≈	504.26
Fudenberg et al. (2012)	658.78	≈	680.53	≈	668.18	<	765.61	≈	779.92	≈	826.75
Kagel and Schley (2013)	2880.11	<	3005.9	≈	3009.28	≈	3032.91	≈	3032.91	≈	3114.31
Sherstyuk et al. (2013)	1472.88	≈	1528.33	≈	1507.33	≈	1536.36	≈	1536.36	<	1624.89
Pooled	**29646.75**	≈	**29874.89**	>	**29578.81**	≪	**31865.03**	≈	**32127.63**	≪	**32913.9**

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 7.

Table 26: 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)	742.43	$\approx$	742.13	$\gg$	694.72	$<$	763.63	$\approx$	763.63	$\approx$	793.71
Blonski et al. (2011)	703.59	$\gg$	585.39	$\gg$	549.45	$\ll$	652.22	$\ll$	691.76	$>$	659.7
Bruttel and Kamecke (2012)	584.62	$\approx$	570.41	$\approx$	567.86	$\approx$	577.25	$\approx$	577.25	$\approx$	594.09
Dal Bó (2005)	376.06	$>$	364.64	$\approx$	358.51	$\approx$	354.58	$\approx$	356.82	$\approx$	365.42
Dal Bó and Fréchette (2011)	3731.81	$\gg$	3594.64	$\approx$	3533.99	$\ll$	3820.5	$\approx$	3827.21	$<$	3954
Dal Bó and Fréchette (2015)	5276.54	$\gg$	5006.42	$\approx$	4991.74	$\ll$	5270.22	$\approx$	5273.98	$\ll$	5396.28
Dreber et al. (2008)	466.07	$>$	451.44	$\approx$	437.17	$\approx$	452.15	$\approx$	452.15	$\approx$	467.69
Duffy and Ochs (2009)	1109.16	$\approx$	1089.19	$\approx$	1090.22	$\approx$	1100.74	$\approx$	1100.74	$\approx$	1119.78
Fréchette and Yuksel (2017)	201.85	$\gg$	164.74	$\approx$	161.45	$\approx$	161.75	$\approx$	164.76	$\approx$	161.75
Fudenberg et al. (2012)	300.67	$\approx$	298.69	$\approx$	291.43	$<$	345.19	$\approx$	345.19	$<$	371.76
Kagel and Schley (2013)	1805.7	$\approx$	1787.59	$\approx$	1782.82	$\approx$	1788.69	$\approx$	1788.69	$<$	1880.45
Sherstyuk et al. (2013)	949.23	$\approx$	924.56	$\approx$	912.8	$<$	958.93	$\approx$	958.93	$\approx$	977.38
Pooled	**16503.06**	$\gg$	**15762.21**	$\gg$	**15481.59**	$\ll$	**16405.97**	$\approx$	**16483.49**	$\ll$	**16851.43**
Second halves per session											
Aoyagi and Frechette (2009)	433.73	$\approx$	433.04	$\gg$	389.24	$<$	458.42	$\approx$	472.12	$\approx$	475.15
Blonski et al. (2011)	982.16	$\gg$	888.46	$\approx$	867.87	$<$	1007.94	$\approx$	1030.01	$>$	1013.07
Bruttel and Kamecke (2012)	340.43	$\approx$	342.01	$\approx$	347.4	$\approx$	340.21	$\approx$	340.21	$\approx$	365
Dal Bó (2005)	431.83	$\approx$	422.93	$\approx$	424.44	$\approx$	442.7	$\approx$	444.23	$\approx$	450.36
Dal Bó and Fréchette (2011)	2882.8	$\approx$	2842.94	$\approx$	2817.31	$\ll$	3124.7	$\approx$	3124.9	$\ll$	3340.71
Dal Bó and Fréchette (2015)	5132.38	$\gg$	5027.27	$\approx$	5043.81	$\ll$	5527	$\approx$	5539.56	$\ll$	5794.56
Dreber et al. (2008)	278.6	$\approx$	271.55	$\approx$	264.94	$\approx$	285.14	$\approx$	285.14	$\approx$	285.64
Duffy and Ochs (2009)	1449.15	$\approx$	1446.73	$>$	1403.03	$\ll$	1587.69	$\approx$	1587.69	$<$	1677.02
Fréchette and Yuksel (2017)	349.28	$\gg$	315.35	$\approx$	313.5	$\approx$	334.45	$\approx$	335.14	$\approx$	334.45
Fudenberg et al. (2012)	394.42	$\approx$	389.34	$\approx$	380.75	$\ll$	440.47	$\approx$	440.47	$<$	488.69
Kagel and Schley (2013)	1204.73	$\approx$	1208.7	$\approx$	1211.37	$\approx$	1240.91	$\approx$	1240.91	$<$	1334.77
Sherstyuk et al. (2013)	587.54	$\approx$	596.62	$\approx$	586.72	$<$	658.71	$\approx$	658.71	$<$	740.5
Pooled	**14722.38**	$\gg$	**14367.31**	$\gg$	**14159.8**	$\ll$	**15620.2**	$\approx$	**15681.45**	$\ll$	**16409.35**

Note: Relation signs, bootstrap procedure, and derived p-values are exactly as above, see Table 1. Pure M1 refers to TFT, Grim, and AD. G2 denotes Grim2. For definitions of the strategies see Table 5.

Table 27: Table 26 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	742.43	≈	742.13	≫	694.72	<	763.63	≈	763.63	≈	793.71
BOS11–9	41.44	≫	29.54	≈	28.34	≪	43.78	≈	46.75	>	43.78
BOS11–14	46.79	≈	45.65	≈	43.74	≪	59.77	<	62.77	≈	59.77
BOS11–15	33.54	≫	17.23	≫	14.27	≪	29.95	≈	32.3	≈	29.95
BOS11–16	114.61	≫	102.32	≈	102.31	≈	105.02	≈	107.37	≈	105.02
BOS11–17	70.92	≫	54.59	≈	48.33	≈	53.64	≈	54.86	≈	53.64
BOS11–26	161.48	≫	136.72	≈	130.46	≈	131.55	≈	131.55	≈	140.64
BOS11–27	50.18	>	46.89	≈	49.15	≈	52.96	≈	57.74	≈	52.96
BOS11–30	36.26	≫	19.4	≫	15.97	≪	30.42	<	33.41	≫	30.42
BOS11–31	78.27	≈	82.99	≈	86.84	≈	113.47	≈	114.93	≈	113.47
BK12–28	584.62	≈	570.41	≈	567.86	≈	577.25	≈	577.25	≈	594.09
D05–18	145.45	≈	140.24	≈	136.21	≈	136.92	≈	138.27	≈	136.92
D05–19	225.65	≈	220.85	≈	220.17	≈	215.01	≈	215.01	≈	226.37
DF11–6	368.33	≫	337.83	≈	333.68	<	379.86	≈	383.44	≈	379.86
DF11–7	542.28	≈	525.23	≈	512.99	≈	543.64	≈	543.64	≈	562.91
DF11–8	565.18	≈	557.31	≈	551.81	<	602.66	≈	602.66	≈	628.99
DF11–22	706.17	≈	685.45	≈	684.04	<	736.15	≈	736.15	≈	748.98
DF11–23	684.75	≈	664.99	≈	656.29	≈	708.88	≈	709.16	≈	763.57
DF11–24	827.4	≈	796.91	≈	779.03	<	820.51	≈	825.22	≈	853.54
DF15–4	192.68	≫	160.51	≈	149.71	≈	158.61	≈	160.16	≈	158.61
DF15–5	792.08	≫	716.06	≈	716.51	≈	740.15	≈	740.15	≈	749.96
DF15–20	973.74	≫	916.66	≈	914.2	≈	933.92	≈	934.42	≈	956.21
DF15–21	1200.7	≈	1201.29	≈	1196.29	≪	1292.16	≈	1292.16	<	1355.35
DF15–33	1775.61	≫	1695.1	≈	1714.32	<	1807.19	≈	1807.19	<	1837.61
DF15–35	300.91	≈	287.64	≈	283.21	<	310.74	≈	310.74	≈	321.04
DRFN08–10	205.38	>	191.06	≈	180.96	≈	193.27	≈	193.27	≈	198.76
DRFN08–11	255.78	≈	256.88	≈	254.11	≈	255.38	≈	255.38	≈	266.82
DO09–32	1109.16	≈	1089.19	≈	1090.22	≈	1100.74	≈	1100.74	≈	1119.78
FY17–25	201.85	≫	164.74	≈	161.45	≈	161.75	≈	164.76	≈	161.75
FRD12–29	300.67	≈	298.69	≈	291.43	<	345.19	≈	345.19	<	371.76
KS13–12	1805.7	≈	1787.59	≈	1782.82	≈	1788.69	≈	1788.69	<	1880.45
STS13–13	949.23	≈	924.56	≈	912.8	<	958.93	≈	958.93	≈	977.38
Aoyagi and Frechette (2009)	742.43	≈	742.13	≫	694.72	<	763.63	≈	763.63	≈	793.71
Blonski et al. (2011)	703.59	≫	585.39	≫	549.45	≪	652.22	≪	691.76	>	659.7
Bruttel and Kamecke (2012)	584.62	≈	570.41	≈	567.86	≈	577.25	≈	577.25	≈	594.09
Dal Bó (2005)	376.06	>	364.64	≈	358.51	≈	354.58	≈	356.82	≈	365.42
Dal Bó and Fréchette (2011)	3731.81	≫	3594.64	≈	3533.99	≪	3820.5	≈	3827.21	<	3954
Dal Bó and Fréchette (2015)	5276.54	≫	5006.42	≈	4991.74	≪	5270.22	≈	5273.98	≪	5396.28
Dreber et al. (2008)	466.07	>	451.44	≈	437.17	≈	452.15	≈	452.15	≈	467.69
Duffy and Ochs (2009)	1109.16	≈	1089.19	≈	1090.22	≈	1100.74	≈	1100.74	≈	1119.78
Fréchette and Yuksel (2017)	201.85	≫	164.74	≈	161.45	≈	161.75	≈	164.76	≈	161.75
Fudenberg et al. (2012)	300.67	≈	298.69	≈	291.43	<	345.19	≈	345.19	<	371.76
Kagel and Schley (2013)	1805.7	≈	1787.59	≈	1782.82	≈	1788.69	≈	1788.69	<	1880.45
Sherstyuk et al. (2013)	949.23	≈	924.56	≈	912.8	<	958.93	≈	958.93	≈	977.38
Pooled	**16503.06**	≫	**15762.21**	≫	**15481.59**	≪	**16405.97**	≈	**16483.49**	≪	**16851.43**

(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	1179.28	≈	1181.85	≫	1085.04	<	1181.51	≈	1224.68	≈	1285.37
BOS11–9	63.01	≫	51.41	≈	50.83	≈	53.75	≈	56.08	≈	53.75
BOS11–14	64.89	≈	63.05	≈	61.74	≪	79.69	≈	82.68	≈	79.69
BOS11–15	43.35	≫	27.05	≫	21.88	≪	39.83	<	42.83	>	39.83
BOS11–16	191.76	>	180.95	≈	178.84	≈	185.23	≈	187.04	≈	185.23
BOS11–17	204.42	>	192.93	≈	183.03	≈	182.96	≈	182.96	≈	190.04
BOS11–26	325.24	≈	319.54	≈	317.33	≈	333.23	≈	333.23	≈	346.08
BOS11–27	211.62	≈	214.41	≈	231.72	≈	231.87	≈	273.97	≈	278.86
BOS11–30	97.9	>	94.64	≈	93.22	≈	100.92	≈	103.29	>	100.92
BOS11–31	182.3	≈	185.95	≈	200.87	≈	200.12	≈	257.66	≈	262.92
BK12–28	917.97	≈	915.39	≈	916.81	≈	925.14	≈	925.14	≈	965.14
D05–18	358.5	≈	356.04	≈	351.89	≈	360.79	≈	360.99	≈	363.86
D05–19	413.11	≈	409.37	≈	418.28	≈	408.94	≈	408.94	<	433.08
DF11–6	600.01	≈	590.39	≈	599.29	<	679.08	≈	683.84	≈	688.5
DF11–7	1198.54	≈	1197.17	≈	1199.27	≈	1254.39	≈	1265.17	≈	1283.74
DF11–8	998.29	≈	996.09	≈	985.67	≪	1081.5	≈	1101.14	≪	1188.89
DF11–22	1338.1	≈	1327.01	≈	1331.85	≈	1428.96	≈	1435.69	<	1505.29
DF11–23	1152.49	≈	1135.01	≈	1113.94	≈	1195.43	≈	1199.3	<	1328.31
DF11–24	1160.83	≈	1159.28	≈	1122.26	<	1184.99	≈	1228.87	≈	1292.35
DF15–4	290.61	≫	271.14	≈	260.02	≈	279.04	≈	286.23	≈	279.04
DF15–5	1374.54	≫	1323.6	≈	1331.49	<	1387.72	≈	1406.13	≈	1436.3
DF15–20	1711.95	≈	1706.95	≈	1716.02	≈	1774.86	≈	1775.13	≈	1843.19
DF15–21	2224.28	≈	2233.11	≈	2213.88	≪	2351.46	≈	2398.7	≪	2559.42
DF15–33	3596.04	≈	3612.75	<	3678.61	<	3913.78	≈	3938.37	<	4057.14
DF15–35	857.8	≈	858.27	≈	856.78	<	972.61	≈	976.66	≈	1044.55
DRFN08–10	287.54	≈	289.36	≈	270.35	<	300.31	≈	303.13	≈	312.98
DRFN08–11	424.73	≈	426.29	≈	425.89	≈	422.42	≈	422.42	≈	435.07
DO09–32	2572.48	≈	2571.83	≈	2526.02	<	2671.36	≈	2708.22	<	2812.9
FY17–25	499.9	>	473.7	≈	473.84	≈	489.74	≈	489.74	≈	489.83
FRD12–29	688.96	≈	680.53	≈	668.18	<	745.36	≈	766.51	<	837.05
KS13–12	3006.47	≈	3005.9	≈	3009.28	≈	3034.78	≈	3037.65	<	3227.41
STS13–13	1532.56	≈	1528.33	≈	1507.33	<	1606.75	≈	1620.49	<	1703.85
Aoyagi and Frechette (2009)	1179.28	≈	1181.85	≫	1085.04	<	1181.51	≈	1224.68	≈	1285.37
Blonski et al. (2011)	1454.6	≫	1380.01	≈	1369.51	<	1448.45	≈	1569.81	≈	1567.35
Bruttel and Kamecke (2012)	917.97	≈	915.39	≈	916.81	≈	925.14	≈	925.14	≈	965.14
Dal Bó (2005)	776.57	≈	768.96	≈	772.3	≈	773.72	≈	773.48	<	799.07
Dal Bó and Fréchette (2011)	6485.95	≈	6431.87	≈	6368.43	≪	6859.43	≈	6940.94	≪	7303.24
Dal Bó and Fréchette (2015)	10096.04	≈	10034.99	≈	10074.29	≪	10710.57	≈	10810.38	≪	11237.13
Dreber et al. (2008)	717.17	≈	719.15	≈	698.35	≈	726.52	≈	729.05	≈	750.15
Duffy and Ochs (2009)	2572.48	≈	2571.83	≈	2526.02	<	2671.36	≈	2708.22	<	2812.9
Fréchette and Yuksel (2017)	499.9	>	473.7	≈	473.84	≈	489.74	≈	489.74	≈	489.83
Fudenberg et al. (2012)	688.96	≈	680.53	≈	668.18	<	745.36	≈	766.51	<	837.05
Kagel and Schley (2013)	3006.47	≈	3005.9	≈	3009.28	≈	3034.78	≈	3037.65	<	3227.41
Sherstyuk et al. (2013)	1532.56	≈	1528.33	≈	1507.33	<	1606.75	≈	1620.49	<	1703.85
Pooled	**30183.29**	≫	**29874.89**	>	**29578.81**	≪	**31380.44**	<	**31778.47**	≪	**33087.89**

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 7.

Table 28: 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"Gene
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)	742.13	$\approx$	740.41	$\approx$	738.78	$\gg$	694.72	$\approx$	740.47	$\approx$	743.27
Blonski et al. (2011)	585.39	$\approx$	570.96	$\gg$	551.67	$\approx$	549.45	$\ll$	681.49	$\ll$	714.05
Bruttel and Kamecke (2012)	570.41	$\approx$	568.94	$\approx$	567.81	$\approx$	567.86	$\approx$	569.25	$\approx$	569.55
Dal Bó (2005)	364.64	$\approx$	360.06	$\approx$	359.02	$\approx$	358.51	$\ll$	395.42	$\ll$	418.7
Dal Bó and Fréchette (2011)	3594.64	$\approx$	3588.23	$\approx$	3577.91	$\approx$	3533.99	$\approx$	3557.12	$\approx$	3576.96
Dal Bó and Fréchette (2015)	5006.42	$\approx$	5011.68	$\approx$	5002.07	$\approx$	4991.74	$\ll$	5225.84	$\approx$	5259.73
Dreber et al. (2008)	451.44	$\approx$	447.65	$\approx$	444.6	$\approx$	437.17	$\approx$	454.47	$\approx$	455.41
Duffy and Ochs (2009)	1089.19	$\approx$	1088.56	$\approx$	1087.11	$\approx$	1090.22	$\approx$	1109.13	$\approx$	1109.16
Fréchette and Yuksel (2017)	164.74	$\approx$	162.94	$\approx$	161.74	$\approx$	161.45	$\approx$	165.9	$\approx$	169.85
Fudenberg et al. (2012)	298.69	$\approx$	300.14	$\approx$	298.27	$\approx$	291.43	$<$	310.29	$\approx$	313.25
Kagel and Schley (2013)	1787.59	$\approx$	1785.44	$\approx$	1783.73	$\approx$	1782.82	$>$	1736.56	$\approx$	1739.6
Sherstyuk et al. (2013)	924.56	$\approx$	925.17	$\approx$	923.27	$\approx$	912.8	$\approx$	916.4	$\approx$	918.43
Pooled	**15762.21**	$\approx$	**15696.08**	$>$	**15605.39**	$>$	**15481.59**	$\ll$	**16044.72**	$\ll$	**16243.29**
Second halves per session											
Aoyagi and Frechette (2009)	433.04	$\approx$	431.24	$\approx$	430.13	$>$	389.24	$<$	436.18	$\approx$	438.02
Blonski et al. (2011)	888.46	$\approx$	889.64	$>$	879.23	$>$	867.87	$<$	962.97	$\approx$	990.65
Bruttel and Kamecke (2012)	342.01	$\approx$	340.72	$\approx$	342.71	$\approx$	347.4	$\approx$	335.98	$\approx$	333.93
Dal Bó (2005)	422.93	$\approx$	419.05	$\approx$	423.8	$\approx$	424.44	$\ll$	479.26	$\approx$	475.76
Dal Bó and Fréchette (2011)	2842.94	$\approx$	2844.4	$\approx$	2835.16	$\approx$	2817.31	$\approx$	2745.37	$\approx$	2759
Dal Bó and Fréchette (2015)	5027.27	$\approx$	5059.55	$\approx$	5058.67	$\approx$	5043.81	$\approx$	5128.18	$\approx$	5109.14
Dreber et al. (2008)	271.55	$\approx$	269.42	$\approx$	266.02	$\approx$	264.94	$\approx$	267.81	$\approx$	271.58
Duffy and Ochs (2009)	1446.73	$\approx$	1444.42	$\approx$	1442.23	$\approx$	1403.03	$\approx$	1461.69	$\approx$	1462.25
Fréchette and Yuksel (2017)	315.35	$\approx$	316.49	$\approx$	314.7	$\approx$	313.5	$\ll$	333.42	$<$	348.67
Fudenberg et al. (2012)	389.34	$\approx$	387.59	$\approx$	385.81	$\approx$	380.75	$\approx$	384.43	$\approx$	386.55
Kagel and Schley (2013)	1208.7	$\approx$	1207.68	$\approx$	1206.72	$\approx$	1211.37	$>$	1156.83	$\approx$	1161.49
Sherstyuk et al. (2013)	596.62	$\approx$	596.25	$\approx$	595.17	$\approx$	586.72	$\approx$	554.97	$\approx$	551.82
Pooled	**14367.31**	$\approx$	**14352.35**	$\approx$	**14289.77**	$>$	**14159.8**	$<$	**14429.45**	$\approx$	**14544.19**

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

Table 29: Table 28 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	742.13	≈	740.41	≈	738.78	≫	694.72	≈	740.47	≈	743.27
BOS11–9	29.54	≈	28.04	≈	28.15	≈	28.34	≪	43.69	≈	44.2
BOS11–14	45.65	≈	44.8	≈	43.39	≈	43.74	≪	56.61	≈	51.98
BOS11–15	17.23	>	15.73	>	14.23	≈	14.27	≪	30.2	≪	33.19
BOS11–16	102.32	≈	101.02	≈	100.59	≈	102.31	<	113.78	≈	114.86
BOS11–17	54.59	≈	53.09	≈	51.86	≈	48.33	≪	68.7	≈	71.21
BOS11–26	136.72	≈	135.26	≈	133.94	≈	130.46	≪	157.04	≈	158.29
BOS11–27	46.89	≈	46.76	≈	46.18	≈	49.15	≈	53.4	≈	55.81
BOS11–30	19.4	≫	17.9	≫	16.4	≈	15.97	≪	33.26	<	36.25
BOS11–31	82.99	≈	88.31	≈	86.87	≈	86.84	≈	74.74	≈	78.15
BK12–28	570.41	≈	568.94	≈	567.81	≈	567.86	≈	569.25	≈	569.55
D05–18	140.24	≈	138.37	≈	136.52	≈	136.21	≪	165.24	≈	161.88
D05–19	220.85	≈	218.85	≈	220.37	≈	220.17	≈	226.64	≪	251.86
DF11–6	337.83	≈	339.19	≈	337.3	≈	333.68	≈	339.43	≈	346.41
DF11–7	525.23	≈	528.44	≈	527.32	≈	512.99	≈	526.64	≈	530.12
DF11–8	557.31	≈	557.06	≈	555.25	≈	551.81	≈	524.93	≈	525.4
DF11–22	685.45	≈	683.66	≈	686.66	≈	684.04	≈	691.45	≈	693.27
DF11–23	664.99	≈	663.28	≈	661.46	≈	656.29	≈	657.42	≈	653.29
DF11–24	796.91	≈	795.06	≈	793.78	≈	779.03	≈	790.32	≈	790.78
DF15–4	160.51	≈	158.66	≈	157.78	≈	149.71	≪	186.25	≈	189.64
DF15–5	716.06	≈	714.63	≈	712.46	≈	716.51	≪	780.89	≈	785.3
DF15–20	916.66	≈	917.24	≈	916.3	≈	914.2	≪	963.41	≈	955.65
DF15–21	1201.29	≈	1198.83	≈	1196.6	≈	1196.29	≈	1211.55	≈	1228.04
DF15–33	1695.1	≈	1711.9	≈	1716.1	≈	1714.32	<	1763.72	≈	1766.98
DF15–35	287.64	≈	287.1	≈	285.35	≈	283.21	≈	290.86	≈	293.3
DRFN08–10	191.06	≈	189.47	≈	187.8	≈	180.96	<	198.75	≈	195.66
DRFN08–11	256.88	≈	255.38	≈	254.69	≈	254.11	≈	252.22	≈	254.85
DO09–32	1089.19	≈	1088.56	≈	1087.11	≈	1090.22	≈	1109.13	≈	1109.16
FY17–25	164.74	≈	162.94	≈	161.74	≈	161.45	≈	165.9	≈	169.85
FRD12–29	298.69	≈	300.14	≈	298.27	≈	291.43	<	310.29	≈	313.25
KS13–12	1787.59	≈	1785.44	≈	1783.73	≈	1782.82	>	1736.56	≈	1739.6
STS13–13	924.56	≈	925.17	≈	923.27	≈	912.8	≈	916.4	≈	918.43
Aoyagi and Frechette (2009)	742.13	≈	740.41	≈	738.78	≫	694.72	≈	740.47	≈	743.27
Blonski et al. (2011)	585.39	≈	570.96	≫	551.67	≈	549.45	≪	681.49	≪	714.05
Bruttel and Kamecke (2012)	570.41	≈	568.94	≈	567.81	≈	567.86	≈	569.25	≈	569.55
Dal Bó (2005)	364.64	≈	360.06	≈	359.02	≈	358.51	≪	395.42	≪	418.7
Dal Bó and Fréchette (2011)	3594.64	≈	3588.23	≈	3577.91	≈	3533.99	≈	3557.12	≈	3576.96
Dal Bó and Fréchette (2015)	5006.42	≈	5011.68	≈	5002.07	≈	4991.74	≪	5225.84	≈	5259.73
Dreber et al. (2008)	451.44	≈	447.65	≈	444.6	≈	437.17	≈	454.47	≈	455.41
Duffy and Ochs (2009)	1089.19	≈	1088.56	≈	1087.11	≈	1090.22	≈	1109.13	≈	1109.16
Fréchette and Yuksel (2017)	164.74	≈	162.94	≈	161.74	≈	161.45	≈	165.9	≈	169.85
Fudenberg et al. (2012)	298.69	≈	300.14	≈	298.27	≈	291.43	<	310.29	≈	313.25
Kagel and Schley (2013)	1787.59	≈	1785.44	≈	1783.73	≈	1782.82	>	1736.56	≈	1739.6
Sherstyuk et al. (2013)	924.56	≈	925.17	≈	923.27	≈	912.8	≈	916.4	≈	918.43
Pooled	**15762.21**	≈	**15696.08**	>	**15605.39**	>	**15481.59**	≪	**16044.72**	≪	**16243.29**

(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	1181.85	≈	1180.06	≈	1178.95	≫	1085.04	<	1179.77	≈	1181.23
BOS11–9	51.41	≈	50.6	≈	50.3	≈	50.83	≪	64.15	≈	62.29
BOS11–14	63.05	≈	62.16	≈	61.57	≈	61.74	≪	78.27	≈	76.73
BOS11–15	27.05	>	25.55	>	24.05	>	21.88	≪	40.08	<	43.07
BOS11–16	180.95	≈	179.45	≈	178.66	≈	178.84	≈	185.48	≈	185.96
BOS11–17	192.93	≈	191.5	≈	190	≈	183.03	<	205.05	≈	195.93
BOS11–26	319.54	≈	317.83	≈	318.19	≈	317.33	≈	331.68	≈	333.84
BOS11–27	214.41	≈	229.34	≈	235.12	≈	231.72	≈	217.95	≈	212.66
BOS11–30	94.64	>	93.32	≈	93.07	≈	93.22	≈	105.76	≈	108.3
BOS11–31	185.95	≈	198.29	≈	200.1	≈	200.87	≈	166.11	≈	165.7
BK12–28	915.39	≈	914.36	≈	915.95	≈	916.81	≈	895.53	≈	891.9
D05–18	356.04	≈	354.4	≈	352.82	≈	351.89	≪	378.37	≈	368.38
D05–19	409.37	≈	407.41	≈	417.01	≈	418.28	≈	420.37	≈	431.08
DF11–6	590.39	≈	596.23	≈	595.49	≈	599.29	≈	579.49	≈	582.07
DF11–7	1197.17	≈	1205.8	≈	1204.41	≈	1199.27	≈	1153.19	≈	1151.03
DF11–8	996.09	≈	999.9	≈	998.05	≈	985.67	≈	915.18	≈	924.46
DF11–22	1327.01	≈	1326.07	≈	1332.1	≈	1331.85	≈	1293.25	≈	1294.96
DF11–23	1135.01	≈	1133.45	≈	1131.71	≈	1113.94	≈	1099.02	≈	1099.93
DF11–24	1159.28	≈	1157.61	≈	1156.21	≈	1122.26	≈	1161.89	≈	1163.33
DF15–4	271.14	≈	269.69	≈	269.21	≈	260.02	≪	296.55	≈	297.48
DF15–5	1323.6	≈	1324.74	≈	1323.34	≈	1331.49	≈	1332.32	≈	1336.61
DF15–20	1706.95	≈	1714.68	≈	1717.74	≈	1716.02	≈	1691.17	≈	1696.55
DF15–21	2233.11	≈	2231.36	≈	2229.41	≈	2213.88	≈	2182.89	≈	2186.44
DF15–33	3612.75	≈	3661.3	≈	3679	≈	3678.61	≈	3682.31	≈	3618.49
DF15–35	858.27	≈	862.75	≈	863.47	≈	856.78	≈	845.23	≈	845.69
DRFN08–10	289.36	≈	287.86	≈	286.23	≈	270.35	≈	292.64	≈	295.95
DRFN08–11	426.29	≈	425.3	≈	424.52	≈	425.89	≈	411.22	≈	413.33
DO09–32	2571.83	≈	2570.1	≈	2568.54	≈	2526.02	≈	2557.19	≈	2559.18
FY17–25	473.7	≈	474.86	≈	473.42	≈	473.84	<	490.2	≈	497.64
FRD12–29	680.53	≈	681.05	≈	679.32	≈	668.18	≈	656.01	≈	658.77
KS13–12	3005.9	≈	3004.83	≈	3004.85	≈	3009.28	≫	2876.28	≈	2880.12
STS13–13	1528.33	≈	1530.53	≈	1529.48	≈	1507.33	≈	1469.2	≈	1472.88
Aoyagi and Frechette (2009)	1181.85	≈	1180.06	≈	1178.95	≫	1085.04	<	1179.77	≈	1181.23
Blonski et al. (2011)	1380.01	≈	1388.1	≈	1381.12	≈	1369.51	≈	1444.6	≈	1454.59
Bruttel and Kamecke (2012)	915.39	≈	914.36	≈	915.95	≈	916.81	≈	895.53	≈	891.9
Dal Bó (2005)	768.96	≈	764.65	≈	771.96	≈	772.3	<	802.29	≈	804.41
Dal Bó and Fréchette (2011)	6431.87	≈	6440.62	≈	6434.12	≈	6368.43	≈	6228.94	≈	6253.48
Dal Bó and Fréchette (2015)	10034.99	≈	10087.85	≈	10099.67	≈	10074.29	≈	10059.64	≈	10022.08
Dreber et al. (2008)	719.15	≈	715.96	≈	712.86	≈	698.35	≈	707.36	≈	714.18
Duffy and Ochs (2009)	2571.83	≈	2570.1	≈	2568.54	≈	2526.02	≈	2557.19	≈	2559.18
Fréchette and Yuksel (2017)	473.7	≈	474.86	≈	473.42	≈	473.84	<	490.2	≈	497.64
Fudenberg et al. (2012)	680.53	≈	681.05	≈	679.32	≈	668.18	≈	656.01	≈	658.77
Kagel and Schley (2013)	3005.9	≈	3004.83	≈	3004.85	≈	3009.28	≫	2876.28	≈	2880.12
Sherstyuk et al. (2013)	1528.33	≈	1530.53	≈	1529.48	≈	1507.33	≈	1469.2	≈	1472.88
Pooled	**29874.89**	≈	**29898.86**	≈	**29859.67**	≫	**29578.81**	≈	**29549.38**	≈	**29645.79**

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 7.

Table 30: 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)	649.81	$\approx$	645.32	$\ll$	791.39	$\approx$	800.02	$\approx$	797.53	$\approx$	799.33
Blonski et al. (2011)	767.03	$\approx$	748.97	$\approx$	705.79	$\approx$	708.13	$\ll$	764.55	$\ll$	795.15
Bruttel and Kamecke (2012)	581.45	$\approx$	617.33	$\approx$	588.54	$\approx$	603.32	$\approx$	590.34	$\approx$	592.12
Dal Bó (2005)	402.78	$\approx$	407.86	$\approx$	389.07	$\approx$	394.44	$\approx$	398.94	$\ll$	414.37
Dal Bó and Fréchette (2011)	3577.56	$\approx$	3624.47	$\ll$	3835.49	$<$	3934.17	$\approx$	3848.71	$\approx$	3866.15
Dal Bó and Fréchette (2015)	5349.43	$\approx$	5355.13	$\ll$	5538.04	$\approx$	5595.6	$\approx$	5573.66	$\approx$	5609.72
Dreber et al. (2008)	487.89	$\approx$	479.28	$>$	462.71	$\approx$	462.71	$\approx$	473.97	$\approx$	479.06
Duffy and Ochs (2009)	1059.59	$\approx$	1056.99	$\approx$	1102.65	$\approx$	1132.23	$\approx$	1106.99	$\approx$	1116.34
Fréchette and Yuksel (2017)	188.15	$\approx$	188.5	$\approx$	181.98	$\approx$	181.98	$\ll$	195.97	$<$	202.98
Fudenberg et al. (2012)	322.89	$\approx$	324.36	$<$	366.75	$\approx$	375.41	$\approx$	368.71	$\approx$	370.66
Kagel and Schley (2013)	1684.63	$<$	1779.94	$\approx$	1805.95	$\approx$	1818.46	$\approx$	1808.32	$\approx$	1812.67
Sherstyuk et al. (2013)	896.24	$\approx$	903.93	$\approx$	941.92	$\approx$	960.83	$\approx$	943.92	$\approx$	946.88
Pooled	**16295.72**	$\approx$	**16350.93**	$\ll$	**16862.01**	$<$	**17076.73**	$\approx$	**17090.45**	$<$	**17260.76**
Second halves per session											
Aoyagi and Frechette (2009)	365.6	$\approx$	363.58	$\ll$	484.4	$\approx$	487.16	$\approx$	490.13	$\approx$	488.8
Blonski et al. (2011)	1022.73	$\approx$	1094.17	$\approx$	1056	$\approx$	1067.5	$\approx$	1108.59	$\approx$	1147.6
Bruttel and Kamecke (2012)	340.71	$\approx$	348.23	$\approx$	316.38	$\approx$	343.47	$\approx$	318.17	$\approx$	319.93
Dal Bó (2005)	471.65	$\approx$	475.11	$\approx$	463.52	$\approx$	464.42	$\approx$	476.02	$\approx$	483.87
Dal Bó and Fréchette (2011)	2706.97	$\ll$	2988.87	$\approx$	2880.23	$\ll$	3251.2	$\gg$	2903.67	$\approx$	2914.93
Dal Bó and Fréchette (2015)	5227.69	$<$	5448.95	$\approx$	5577.08	$\ll$	5853.69	$\gg$	5603.14	$\approx$	5635.29
Dreber et al. (2008)	288.95	$\approx$	295.06	$\approx$	287.58	$\approx$	287.58	$<$	298.37	$\approx$	301.17
Duffy and Ochs (2009)	1370.98	$\approx$	1409.56	$\ll$	1617.77	$\approx$	1661.56	$\approx$	1620.05	$\approx$	1622.69
Fréchette and Yuksel (2017)	311.99	$\approx$	309.63	$\ll$	356.11	$\approx$	356.11	$\approx$	362.81	$\approx$	364.77
Fudenberg et al. (2012)	364.47	$\approx$	373.44	$<$	447.19	$\approx$	476.66	$>$	449.08	$\approx$	449.92
Kagel and Schley (2013)	1128.24	$\approx$	1211.25	$\approx$	1169.31	$<$	1274.78	$>$	1171.67	$\approx$	1172
Sherstyuk et al. (2013)	551.88	$<$	616.19	$\approx$	583.8	$\ll$	691.06	$\gg$	585.8	$\approx$	587.81
Pooled	**14480.14**	$\ll$	**15152.88**	$\approx$	**15396.72**	$\ll$	**16324.6**	$\gg$	**15606.38**	$\approx$	**15744.12**

Note: Relation signs, bootstrap procedure, and derived p-values are exactly as above, see Table 1. Pure M1 refers to TFT, Grim, and AD. G2 denotes Grim2. For definitions of pure strategies see Table 5. 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 31: Table 30 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	649.81	≈	645.32	≪	791.39	≈	800.02	≈	797.53	≈	799.33
BOS11–9	42.83	≈	39.68	≪	47.66	≈	47.66	≪	52.51	≈	54.13
BOS11–14	68.28	≈	72.3	≫	61.41	≈	61.41	<	65.9	≈	67.38
BOS11–15	45.19	≈	42.36	≫	29.87	≈	29.87	≪	34.37	≈	35.93
BOS11–16	110.81	≈	106.93	≈	109.88	≈	109.88	≈	114.37	≈	118.13
BOS11–17	56.33	≈	56.31	≫	48.65	≈	48.65	≈	52.5	≈	53.99
BOS11–26	164.22	≈	168.38	≈	162.85	≈	162.85	≈	163.53	<	167.5
BOS11–27	57.51	<	71.67	>	61.4	≈	67.19	≈	61.4	≈	62.62
BOS11–30	47.12	≫	42.87	≫	30.39	≈	30.39	≪	35.13	≈	36.68
BOS11–31	84.61	≈	88.38	≈	120.17	≈	120.17	≈	124.75	≈	128.69
BK12–28	581.45	≈	617.33	≈	588.54	≈	603.32	≈	590.34	≈	592.12
D05–18	161.9	≈	154.75	≈	153.78	≈	153.78	≈	159.8	<	165.13
D05–19	234.5	≈	248.85	>	232.63	≈	238.53	≈	234.88	<	244.28
DF11–6	335.02	≈	339.47	≈	368.99	≈	389.67	≈	370.93	≈	372.82
DF11–7	535.47	≈	526.77	<	565.34	≈	566.75	≈	567.52	≈	569.45
DF11–8	553.44	<	619.1	≈	568.74	<	613.19	≈	569.64	≈	569.91
DF11–22	693.74	≈	707.69	<	756.01	≈	758.47	≈	757.9	≈	761.76
DF11–23	641.88	≈	658.75	<	718.87	≈	746.74	≈	720.7	≈	722.73
DF11–24	769.56	≈	740.38	<	829.71	≈	843.2	≈	829.71	≈	831.79
DF15–4	177.53	≈	175.94	≈	170.85	≈	170.85	<	176.64	<	183.07
DF15–5	738.21	≈	737.26	≈	772.27	≈	775.66	≈	774.65	≈	777.35
DF15–20	952.67	≈	960.95	≈	999.39	≈	1005.78	≈	1002.08	≈	1012.09
DF15–21	1268.05	≈	1269	<	1335.48	≈	1389.45	>	1337.99	≈	1341.36
DF15–33	1857.19	≈	1865.4	≈	1900.52	≈	1900.52	≈	1909.83	≈	1915.81
DF15–35	303.29	≈	311.6	≈	334.35	≈	335.83	≈	337.48	≈	339.22
DRFN08–10	215.08	≈	210.6	>	200.93	≈	200.93	≈	205.93	≈	207.6
DRFN08–11	266.5	≈	264.48	≈	259.68	≈	259.68	≈	263.84	≈	266.56
DO09–32	1059.59	≈	1056.99	≈	1102.65	≈	1132.23	≈	1106.99	≈	1116.34
FY17–25	188.15	≈	188.5	≈	181.98	≈	181.98	≪	195.97	<	202.98
FRD12–29	322.89	≈	324.36	<	366.75	≈	375.41	≈	368.71	≈	370.66
KS13–12	1684.63	<	1779.94	≈	1805.95	≈	1818.46	≈	1808.32	≈	1812.67
STS13–13	896.24	≈	903.93	≈	941.92	≈	960.83	≈	943.92	≈	946.88
Aoyagi and Frechette (2009)	649.81	≈	645.32	≪	791.39	≈	800.02	≈	797.53	≈	799.33
Blonski et al. (2011)	767.03	≈	748.97	≈	705.79	≈	708.13	≪	764.55	≪	795.15
Bruttel and Kamecke (2012)	581.45	≈	617.33	≈	588.54	≈	603.32	≈	590.34	≈	592.12
Dal Bó (2005)	402.78	≈	407.86	≈	389.07	≈	394.44	≈	398.94	≪	414.37
Dal Bó and Fréchette (2011)	3577.56	≈	3624.47	≪	3835.49	<	3934.17	≈	3848.71	≈	3866.15
Dal Bó and Fréchette (2015)	5349.43	≈	5355.13	≪	5538.04	≈	5595.6	≈	5573.66	≈	5609.72
Dreber et al. (2008)	487.89	≈	479.28	>	462.71	≈	462.71	≈	473.97	≈	479.06
Duffy and Ochs (2009)	1059.59	≈	1056.99	≈	1102.65	≈	1132.23	≈	1106.99	≈	1116.34
Fréchette and Yuksel (2017)	188.15	≈	188.5	≈	181.98	≈	181.98	≪	195.97	<	202.98
Fudenberg et al. (2012)	322.89	≈	324.36	<	366.75	≈	375.41	≈	368.71	≈	370.66
Kagel and Schley (2013)	1684.63	<	1779.94	≈	1805.95	≈	1818.46	≈	1808.32	≈	1812.67
Sherstyuk et al. (2013)	896.24	≈	903.93	≈	941.92	≈	960.83	≈	943.92	≈	946.88
Pooled	**16295.72**	≈	**16350.93**	≪	**16862.01**	<	**17076.73**	≈	**17090.45**	<	**17260.76**

(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	1002.43	≈	999.39	≪	1257.49	≈	1294.67	>	1258.18	≈	1257.49
BOS11–9	65.74	≈	56.83	>	52.33	≈	52.33	<	56.83	≈	58.33
BOS11–14	76.34	≈	75.9	≈	82.94	≈	82.94	<	87.44	≈	88.94
BOS11–15	54.87	>	52.15	≫	39.65	≈	39.65	≪	44.14	≈	45.76
BOS11–16	176.37	≈	170.31	≈	179.09	≈	179.09	≈	183.58	≈	185.1
BOS11–17	198.2	≈	194.51	≈	191.8	≈	191.8	≈	193.3	≈	194.79
BOS11–26	336.31	<	366.5	≈	356.53	≈	356.53	≈	362.07	≈	366.54
BOS11–27	220.67	≈	276.32	≈	264.28	≈	274.29	≈	265.78	≈	267.39
BOS11–30	105.65	≈	116.18	>	110.41	≈	110.41	≈	114.91	≈	116.4
BOS11–31	183.03	≈	268.96	≈	258.45	≈	264.93	≈	259.95	≈	261.45
BK12–28	911.92	≈	948.27	≈	912.09	≈	954.26	≈	913.88	≈	915.75
D05–18	359.61	≈	360.73	≈	365.93	≈	365.93	≈	373.59	≈	378.69
D05–19	423.98	≈	442.98	≈	434.95	≈	440.25	≈	437.41	≈	441.34
DF11–6	556.98	≈	584.49	≈	633.71	≈	699.65	≈	635.6	≈	637.42
DF11–7	1126.45	≈	1135.74	≈	1202.27	≈	1242.47	≈	1204.37	≈	1206.28
DF11–8	986.17	≈	1007.13	≈	1076.28	<	1131.38	≈	1078.2	≈	1079.16
DF11–22	1355.57	<	1469.16	≈	1448.29	≈	1488.69	≈	1451.35	≈	1448.29
DF11–23	1068.94	≈	1105.85	≈	1185.24	≈	1294.05	>	1193.61	≈	1185.24
DF11–24	1080.01	≈	1083.47	<	1259	≈	1296.29	≈	1259	≈	1261.46
DF15–4	286.93	≈	298.69	≈	299.85	≈	299.85	≈	305.71	≈	309.31
DF15–5	1350.25	≈	1350.84	<	1428.84	≈	1445.69	≈	1431.31	≈	1441.57
DF15–20	1720.89	≈	1770.03	≈	1813.65	≈	1859.17	≈	1816.08	≈	1823.16
DF15–21	2240.55	<	2345.14	≈	2421.02	<	2536.56	>	2423.58	≈	2432.56
DF15–33	3798.6	≪	3964.95	<	4103.22	≈	4132.33	≈	4103.22	≈	4106.97
DF15–35	907.25	≈	1020.65	≈	1007.18	≈	1060.46	>	1008.97	≈	1010.76
DRFN08–10	315.81	≈	314.76	≈	313.57	≈	313.57	≈	318.57	≈	320.24
DRFN08–11	425.55	≈	419.57	≈	436.15	≈	436.15	≈	440.79	≈	436.94
DO09–32	2428.38	≈	2440.62	≪	2780.93	≈	2823.45	≈	2787.08	≈	2783.33
FY17–25	480.44	≈	475.52	≈	504.26	≈	504.26	≈	510.95	≈	512.88
FRD12–29	664.6	≈	683.14	<	779.91	≈	826.75	>	781.86	≈	786.16
KS13–12	2871.48	<	3030.1	≈	3032.91	≈	3114.32	≈	3035.28	≈	3040.47
STS13–13	1426.09	≈	1514.33	≈	1536.36	<	1624.89	≈	1538.37	≈	1540.4
Aoyagi and Frechette (2009)	1002.43	≈	999.39	≪	1257.49	≈	1294.67	>	1258.18	≈	1257.49
Blonski et al. (2011)	1507.31	≈	1637.77	>	1570.14	≈	1582.03	≈	1628.08	≈	1654.79
Bruttel and Kamecke (2012)	911.92	≈	948.27	≈	912.09	≈	954.26	≈	913.88	≈	915.75
Dal Bó (2005)	789.96	≈	807.96	≈	803.54	≈	808.31	≈	815.25	≈	824.99
Dal Bó and Fréchette (2011)	6222.59	<	6418.15	≪	6836.38	≪	7168.69	>	6854.44	≈	6855.56
Dal Bó and Fréchette (2015)	10356.95	≪	10785.28	<	11101	<	11351.56	≫	11123.87	≈	11165.17
Dreber et al. (2008)	747.66	≈	738.54	≈	751.82	≈	751.82	≈	763.56	≈	762.08
Duffy and Ochs (2009)	2428.38	≈	2440.62	≪	2780.93	≈	2823.45	≈	2787.08	≈	2783.33
Fréchette and Yuksel (2017)	480.44	≈	475.52	≈	504.26	≈	504.26	≈	510.95	≈	512.88
Fudenberg et al. (2012)	664.6	≈	683.14	<	779.91	≈	826.75	>	781.86	≈	786.16
Kagel and Schley (2013)	2871.48	<	3030.1	≈	3032.91	≈	3114.32	≈	3035.28	≈	3040.47
Sherstyuk et al. (2013)	1426.09	≈	1514.33	≈	1536.36	<	1624.89	≈	1538.37	≈	1540.4
Pooled	**29738.08**	≪	**30697.93**	≪	**32032.91**	≪	**32914.44**	≫	**32229.66**	≈	**32354.4**

Note: Notation of treatments and meaning of relation signs are all as defined above, see Table 7.

Table 32: 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)	742.13	$\approx$	738.78	$\gg$	694.72	$>$	649.81	$\approx$	645.32	$\ll$	791.39
Blonski et al. (2011)	585.39	$\gg$	551.67	$\approx$	549.45	$\ll$	767.03	$\approx$	748.97	$\approx$	705.79
Bruttel and Kamecke (2012)	570.41	$\approx$	567.81	$\approx$	567.86	$\approx$	581.45	$\approx$	617.33	$\approx$	588.54
Dal Bó (2005)	364.64	$\approx$	359.02	$\approx$	358.51	$<$	402.78	$\approx$	407.86	$\approx$	389.07
Dal Bó and Fréchette (2011)	3594.64	$\approx$	3577.91	$\approx$	3533.99	$\approx$	3577.56	$\approx$	3624.47	$\ll$	3835.49
Dal Bó and Fréchette (2015)	5006.42	$\approx$	5002.07	$\approx$	4991.74	$\ll$	5349.43	$\approx$	5355.13	$\ll$	5538.04
Dreber et al. (2008)	451.44	$\approx$	444.6	$\approx$	437.17	$<$	487.89	$\approx$	479.28	$>$	462.71
Duffy and Ochs (2009)	1089.19	$\approx$	1087.11	$\approx$	1090.22	$\approx$	1059.59	$\approx$	1056.99	$\approx$	1102.65
Fréchette and Yuksel (2017)	164.74	$\approx$	161.74	$\approx$	161.45	$\ll$	188.15	$\approx$	188.5	$\approx$	181.98
Fudenberg et al. (2012)	298.69	$\approx$	298.27	$\approx$	291.43	$<$	322.89	$\approx$	324.36	$<$	366.75
Kagel and Schley (2013)	1787.59	$\approx$	1783.73	$\approx$	1782.82	$>$	1684.63	$<$	1779.94	$\approx$	1805.95
Sherstyuk et al. (2013)	924.56	$\approx$	923.27	$\approx$	912.8	$\approx$	896.24	$\approx$	903.93	$\approx$	941.92
Pooled	**15762.21**	$\gg$	**15605.39**	$>$	**15481.59**	$\ll$	**16295.72**	$\approx$	**16350.93**	$\ll$	**16862.01**
Second halves per session											
Aoyagi and Frechette (2009)	433.04	$\approx$	430.13	$>$	389.24	$\approx$	365.6	$\approx$	363.58	$\ll$	484.4
Blonski et al. (2011)	888.46	$\approx$	879.23	$>$	867.87	$\ll$	1022.73	$\approx$	1094.17	$\approx$	1056
Bruttel and Kamecke (2012)	342.01	$\approx$	342.71	$\approx$	347.4	$\approx$	340.71	$\approx$	348.23	$\approx$	316.38
Dal Bó (2005)	422.93	$\approx$	423.8	$\approx$	424.44	$\ll$	471.65	$\approx$	475.11	$\approx$	463.52
Dal Bó and Fréchette (2011)	2842.94	$\approx$	2835.16	$\approx$	2817.31	$\approx$	2706.97	$\ll$	2988.87	$\approx$	2880.23
Dal Bó and Fréchette (2015)	5027.27	$\approx$	5058.67	$\approx$	5043.81	$<$	5227.69	$<$	5448.95	$\approx$	5577.08
Dreber et al. (2008)	271.55	$\approx$	266.02	$\approx$	264.94	$\approx$	288.95	$\approx$	295.06	$\approx$	287.58
Duffy and Ochs (2009)	1446.73	$\approx$	1442.23	$\approx$	1403.03	$\approx$	1370.98	$\approx$	1409.56	$\ll$	1617.77
Fréchette and Yuksel (2017)	315.35	$\approx$	314.7	$\approx$	313.5	$\approx$	311.99	$\approx$	309.63	$\ll$	356.11
Fudenberg et al. (2012)	389.34	$\approx$	385.81	$\approx$	380.75	$\approx$	364.47	$\approx$	373.44	$<$	447.19
Kagel and Schley (2013)	1208.7	$\approx$	1206.72	$\approx$	1211.37	$>$	1128.24	$\approx$	1211.25	$\approx$	1169.31
Sherstyuk et al. (2013)	596.62	$\approx$	595.17	$\approx$	586.72	$\approx$	551.88	$<$	616.19	$\approx$	583.8
Pooled	**14367.31**	$\approx$	**14289.77**	$>$	**14159.8**	$\approx$	**14480.14**	$\ll$	**15152.88**	$\approx$	**15396.72**

Note: Relation signs, bootstrap procedure, and derived p-values are exactly as above, see Table 1. Pure M1 refers to TFT, Grim, and AD. For definitions of pure strategies see Table 5. 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 33: 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)	742.13	$\approx$	738.55	$\gg$	694.72	$>$	649.37	$\approx$	645.32	$\ll$	791.38
Blonski et al. (2011)	585.39	$\gg$	550.9	$\approx$	549.45	$\ll$	760.49	$\approx$	748.97	$\approx$	703.1
Bruttel and Kamecke (2012)	570.41	$\approx$	567.91	$\approx$	567.86	$\approx$	580.92	$\approx$	617.33	$\approx$	588.55
Dal Bó (2005)	364.64	$\approx$	359.74	$\approx$	358.51	$\ll$	405.83	$\approx$	407.86	$\approx$	389.08
Dal Bó and Fréchette (2011)	3594.64	$\approx$	3573.88	$\approx$	3533.99	$\approx$	3524.12	$<$	3624.47	$\ll$	3835.75
Dal Bó and Fréchette (2015)	5006.42	$\approx$	5004.6	$\approx$	4991.74	$\ll$	5276.75	$\approx$	5355.13	$\ll$	5538.37
Dreber et al. (2008)	451.44	$\approx$	445.38	$\approx$	437.17	$<$	482.83	$\approx$	479.28	$>$	462.71
Duffy and Ochs (2009)	1089.19	$\approx$	1090.75	$\approx$	1090.22	$\approx$	1039.25	$\approx$	1056.99	$\approx$	1102.64
Fréchette and Yuksel (2017)	164.74	$\approx$	161.91	$\approx$	161.45	$\ll$	182.81	$\approx$	188.5	$\approx$	181.98
Fudenberg et al. (2012)	298.69	$\approx$	296.73	$\approx$	291.43	$<$	319.76	$\approx$	324.36	$<$	366.78
Kagel and Schley (2013)	1787.59	$\approx$	1784.29	$\approx$	1782.82	$>$	1651.6	$\ll$	1779.94	$\approx$	1805.95
Sherstyuk et al. (2013)	924.56	$\approx$	922.74	$\approx$	912.8	$\approx$	890.54	$\approx$	903.93	$\approx$	941.91
Pooled	**15762.21**	$\gg$	**15606.79**	$>$	**15481.59**	$\ll$	**16092.54**	$<$	**16350.93**	$\ll$	**16858.86**
Second halves per session											
Aoyagi and Frechette (2009)	433.04	$\approx$	431.45	$\gg$	389.24	$\approx$	368.09	$\approx$	363.58	$\ll$	484.41
Blonski et al. (2011)	888.46	$\approx$	874.57	$\approx$	867.87	$\ll$	1013.57	$\approx$	1094.17	$\approx$	1055.95
Bruttel and Kamecke (2012)	342.01	$\approx$	345.65	$\approx$	347.4	$\approx$	324.09	$\approx$	348.23	$\approx$	316.38
Dal Bó (2005)	422.93	$\approx$	424.63	$\approx$	424.44	$<$	451.34	$<$	475.11	$\approx$	463.54
Dal Bó and Fréchette (2011)	2842.94	$\approx$	2833.79	$\approx$	2817.31	$\approx$	2679.96	$\ll$	2988.87	$\approx$	2885.43
Dal Bó and Fréchette (2015)	5027.27	$\approx$	5048.25	$\approx$	5043.81	$\approx$	5144.02	$\ll$	5448.95	$\approx$	5577.55
Dreber et al. (2008)	271.55	$\approx$	270.81	$\approx$	264.94	$\approx$	287.55	$\approx$	295.06	$\approx$	287.58
Duffy and Ochs (2009)	1446.73	$\approx$	1442.34	$\approx$	1403.03	$\approx$	1349.18	$<$	1409.56	$\ll$	1617.77
Fréchette and Yuksel (2017)	315.35	$\approx$	312.94	$\approx$	313.5	$\approx$	311.04	$\approx$	309.63	$\ll$	356.11
Fudenberg et al. (2012)	389.34	$\approx$	386.5	$\approx$	380.75	$\approx$	367.2	$\approx$	373.44	$<$	447.19
Kagel and Schley (2013)	1208.7	$\approx$	1208.4	$\approx$	1211.37	$>$	1100.94	$\approx$	1211.25	$\approx$	1169.31
Sherstyuk et al. (2013)	596.62	$\approx$	595.53	$\approx$	586.72	$>$	541.54	$<$	616.19	$\approx$	583.8
Pooled	**14367.31**	$\approx$	**14284.29**	$>$	**14159.8**	$\approx$	**14266.78**	$\ll$	**15152.88**	$\approx$	**15402.36**

Note: Relation signs, bootstrap procedure, and derived p-values are exactly as above, see Table 1. Pure M1 refers to TFT, Grim, and AD. For definitions of pure strategies see Table 5. "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 34: Examining all mixtures of Semi-Grim with pure or generalized pure strategies as secondary components (robustness check for Table 4)

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)	684.3	$\approx$	689.91	$\approx$	688.17	$\approx$	649.86	$\approx$	698.36	$\approx$	681.99	$\approx$	698.37	$\approx$	698.36
Blonski et al. (2011)	635.26	$\approx$	648.6	$\approx$	645.78	$\gg$	614.35	$\ll$	663.28	$\approx$	658.75	$>$	614.21	$>$	594.32
Bruttel and Kamecke (2012)	572.36	$\approx$	559.64	$\approx$	571.61	$\approx$	562.11	$\approx$	567.84	$\approx$	562.35	$\approx$	571.1	$\approx$	566.52
Dal Bó (2005)	393.92	$\approx$	372.81	$\approx$	384.74	$\approx$	372.89	$\approx$	378.87	$\approx$	389.65	$\approx$	386.98	$>$	368.21
Dal Bó and Fréchette (2011)	3538.98	$\approx$	3532.69	$\approx$	3475.81	$\approx$	3416.01	$\approx$	3488.81	$\approx$	3434.87	$<$	3538.27	$\approx$	3577.29
Dal Bó and Fréchette (2015)	5164.89	$\gg$	5070.81	$\approx$	5052.96	$\gg$	4941.84	$\approx$	5024.63	$\approx$	5107.87	$>$	5047.74	$\approx$	5033.06
Dreber et al. (2008)	454.45	$\approx$	453.19	$\approx$	440.94	$\approx$	443.12	$\approx$	442.47	$\approx$	440.24	$\approx$	443.9	$\approx$	444.99
Duffy and Ochs (2009)	1084.07	$\approx$	1062.36	$\approx$	1074.19	$>$	1017.65	$<$	1084.74	$\approx$	1074.22	$\approx$	1099.57	$\approx$	1084.81
Fréchette and Yuksel (2017)	180.17	$\approx$	190.03	$\approx$	188.6	$\gg$	167.94	$\approx$	165.94	$<$	178.18	$\approx$	180.54	$\gg$	165.36
Fudenberg et al. (2012)	296.17	$\approx$	299.94	$\approx$	299.98	$\approx$	291.68	$\approx$	292.47	$\approx$	295.53	$\approx$	297.99	$\approx$	295.46
Kagel and Schley (2013)	1785.45	$\approx$	1780	$\approx$	1777.41	$>$	1690.55	$\approx$	1689.17	$\approx$	1685.56	$<$	1777.58	$\approx$	1787.56
Sherstyuk et al. (2013)	910.79	$\approx$	901.32	$\approx$	911	$>$	866.46	$\approx$	887.79	$\approx$	889.32	$\approx$	904.29	$\approx$	916.83
Pooled	**15919.66**	$>$	**15780.15**	$\approx$	**15730.04**	$\gg$	**15253.3**	$\ll$	**15566.74**	$\approx$	**15580.91**	$\approx$	**15742.91**	$\approx$	**15715.14**
Second halves per session															
Aoyagi and Frechette (2009)	395.39	$\approx$	380.72	$\approx$	403.19	$>$	357.21	$\approx$	392.87	$\approx$	398.09	$\approx$	397.94	$\approx$	393.02
Blonski et al. (2011)	921.84	$\approx$	946.43	$<$	1157.9	$\gg$	881.73	$\approx$	914.98	$\approx$	915.1	$\approx$	940.75	$>$	919.01
Bruttel and Kamecke (2012)	323.04	$\approx$	339	$\approx$	337.37	$\approx$	346.24	$\approx$	349.32	$\approx$	336.68	$\approx$	337.22	$\approx$	350.98
Dal Bó (2005)	471.52	$\approx$	461.35	$\approx$	444.45	$\approx$	437.05	$\approx$	437.71	$<$	457.32	$\approx$	444.73	$\approx$	433.69
Dal Bó and Fréchette (2011)	2729.84	$\approx$	2742.23	$\ll$	3652.93	$\gg$	2620.29	$\approx$	2640.88	$\approx$	2615.25	$<$	2798.46	$\approx$	2845.99
Dal Bó and Fréchette (2015)	5043.7	$\approx$	5080.21	$\approx$	5025.85	$>$	4925.77	$\approx$	5032.59	$\approx$	5077.82	$\approx$	5053.74	$\approx$	5042.75
Dreber et al. (2008)	279.59	$\approx$	258.73	$\approx$	276.69	$\approx$	279.28	$\approx$	275.84	$\approx$	277.77	$\approx$	262.93	$\approx$	275.37
Duffy and Ochs (2009)	1407.13	$\approx$	1378.8	$\approx$	1365.2	$>$	1294.4	$<$	1399.65	$\approx$	1379.73	$\approx$	1414.84	$\approx$	1447.23
Fréchette and Yuksel (2017)	313.59	$\approx$	324.17	$\approx$	323.61	$>$	283.31	$<$	319.78	$\approx$	311.64	$\approx$	319.44	$\approx$	317.41
Fudenberg et al. (2012)	381.7	$\approx$	387.98	$\approx$	385.9	$\approx$	364.37	$\approx$	366.79	$\approx$	379.95	$\approx$	388.9	$\approx$	384.62
Kagel and Schley (2013)	1156.16	$\approx$	1202.31	$\approx$	1202.28	$\approx$	1154.97	$\approx$	1126.89	$\approx$	1153.79	$\approx$	1200.87	$\approx$	1216.1
Sherstyuk et al. (2013)	588.67	$\approx$	569.49	$\approx$	583.54	$\approx$	565.48	$\approx$	550.64	$\approx$	539.35	$\approx$	567.48	$\approx$	594.92
Pooled	**14231.02**	$\approx$	**14290.27**	$\ll$	**15377.76**	$\gg$	**13728.94**	$<$	**13990.32**	$\approx$	**14024.87**	$<$	**14309.69**	$\approx$	**14403.47**

Note: Relation signs, bootstrap procedure, and derived p-values are exactly as above, see Table 1. For definitions of pure strategies see Table 5. For definitions of generalized strategies see Section 3 main text.