Predicting Cooperation with Learning Models*

Drew Fudenberg† Gustav Karreskog ‡

First posted version: October 21, 2020
This version: November 16, 2021

Abstract

We use simulations of a simple learning model to predict how cooperation varies with treatment in the experimental play of the indefinitely repeated prisoner's dilemma. We suppose that learning and the game parameters only influence play in the initial round of each supergame, and that after these rounds play depends only on the outcome of the previous round. Using data from 17 papers, we find that our model predicts out-of-sample cooperation at least as well as more complicated models with more parameters and harder-to-interpret machine learning algorithms. Our results let us predict how cooperation rates change with longer experimental sessions, and help explain past findings on the role of strategic uncertainty.

Keywords: cooperation, prisoner's dilemma, risk dominance, predictive game theory

JEL codes: C53, C63, C72, C92, D83, D9

*We thank Anna Dreber Almenberg, Mathias Blonski, Yves Breitmoser, Olivier Compte, Glenn Ellison, Ying Gao, Steven Lehrer, Annie Liang, Erik Mohlin, Indira Puri, Karl Schlag, Emanuel Vespa, Jörgen Weibull, David Rand, Alex Wolitzky, and seminar participants at Goethe University Frankfurt, the Harvard-MIT theory workshop, SITE, U Queensland, UVA, and VIBES for helpful comments. NSF grants SES 1643517 and 1951056, the Jan Wallander and Tom Hedelius Foundation, and the Knut and Alice Wallenberg Research Foundation provided financial support.

†Department of Economics, MIT, 77 Massachusetts Avenue, Cambridge, MA, 02139; drew.fudenberg@gmail.com

‡Department of Economics, Uppsala University, Kyrkogårdsgatan 10, 751 20, Uppsala, Sweden; gustav.karreskog@nek.uu.se

1 Introduction

Determining when and how people overcome short-run incentives to behave cooperatively is a key issue in the social sciences. The theory of repeated games has determined which factors allow cooperation as an equilibrium outcome, but since these games typically also have equilibria where people do not cooperate, equilibrium theory on its own is not a useful way of making predictions about cooperation rates. Moreover, the assumption that people play the most cooperative equilibrium possible, which is often used in applications, is a very poor fit for observed behavior in the laboratory. It is therefore important both for policy decisions and the development of more useful theories to have a better understanding of how cooperation rates in experimental play of repeated games depend on their parameters.

To that end, we treat the relation between the average cooperation rate in the experimental play of the prisoner's dilemma and its exogenous parameters as a prediction problem. That is, in contrast to past work, which has tried to match observed behavior in-sample by, e.g., conditioning on initial play or entire sequence of previous actions, we try to predict the average cooperation in a session without using any data about how people in the session actually played. Studying predictions rather than in-sample fit helps us identify aspects of behavior that are stable across experiments and significant determinants of behavior.

To make our predictions, we use a very simple model of reinforcement learning, where all that varies with treatment or personal experience is the probability of cooperating in the first round of a new match. After these initial rounds, play depends only on the outcome of the previous round: If both players cooperated they keep cooperating, if they both defected they keep defecting, and if they mismatch, i.e., one player cooperated and one defected, they both cooperate with roughly 1/3 probability. Of course, actual behavior is much more complicated than our model supposes, and exhibit many individual-specific effects. However, our goal here is not to identify all aspects that *sometimes* matter, and include them in a model. Such a model would be difficult to interpret, and could only be estimated with orders of magnitude more data than is currently available.

In our learning model, the way that people play in the first round of their very first supergame depends on a composite parameter Δ^{RD} that captures some of the effect of strategic uncertainty. This parameter is defined as the difference between

the actual discount factor of the game and the discount factor that makes players indifferent between the strategies Grim and Always Defect on the assumption that everyone in the population uses one of these strategies, and moreover, that exactly half of the population uses each one. (This is not meant as a realistic assumption, but is simply one explanation of how Δ^{RD} is defined.) The initial choices in a supergame and the fixed strategy in subsequent rounds of the supergame determine the payoffs in that supergame. Initial-round play in following supergames depends on Δ^{RD} and on the overall payoffs the player received in past supergames for each initial-round action. Because of this feedback channel, the predicted overall cooperation rates end up depending not only on the composite parameter Δ^{RD}, but also on the individual game parameters.

To make outsample predictions, we use simulations of play, not endogenous data such as the actions played and the payoffs received. We find the parameters that best fit the time paths of cooperation on our training sets, and evaluate the predictions that the parameters generate on test sets. Using data from the 103 experimental sessions gathered in Dal Bó and Fréchette (2018) as well as 58 sessions in papers published since then, we find that the learning model predicts both average cooperation and the time path of cooperation in a session at least as well as any of the black-box methods we consider, and better than alternative learning models based on pure strategies. Moreover, we find that allowing for heterogeneous agents, or a more complex learning model with learning at all memory-1 histories, gives no noticeable improvements.

The learning model also allows us to predict what would happen in longer experimental sessions than are typical in the lab. As a first step, we use data from the first halves of some laboratory sessions to predict cooperation rates in the second halves of other sessions, and find that our model has significantly better outsample predictions than OLS on Δ^{RD}, Lasso, support vector regression (SVR), or a gradient boosting tree (GBT). We then use the model to simulate cooperation rates in extremely long sessions. We find that, as expected, cooperation rates are low when Δ^{RD} is negative and that cooperation rates are high when Δ^{RD} is large. However, for intermediate values of Δ^{RD} there is substantial cross-session variability in cooperation even in the long run: Even after 10,000 supergames, the 90% interval goes from 0% to 79%.

As we detail in Section 3, past work has already found evidence that overall cooperation rates depend on Δ^{RD}. Our preliminary data analysis sharpens this conclusion: cooperation tends to increase over the course of a session when $\Delta^{RD} > 0.15$,

and to decrease when $\Delta^{RD} < 0$. Our learning model predicts this pattern, which suggests that the reason for the observed impact of the composite parameter Δ^{RD} is its effect on the reinforcement of cooperation in the initial round of each supergame. Furthermore, our model predicts that he individual parameters matter in addition to the way their effect on Δ^{RD}. This helps explain why our learning model can make better predictions than models using Δ^{RD} alone.

Our model also predicts that participants who use the same learning rule can encounter different behavior by their partners because of their own randomization in the initial rounds and the random matching of partners. This difference in outcomes can lead participants with the same learning rule to behave very differently, and we argue that ignoring this endogenous heterogeneity may lead to overestimates of how different the participants are ex-ante. For example, our reinforcement learning model can explain why a sizable number participants consistently defect in treatments that are very conducive to cooperation. Finally, our learning model replicates the empirical regularity that there is more cross-session variation in average cooperation rates for intermediate values of Δ^{RD}. Among other things, this suggests caution in interpreting experimental studies of prisoner's dilemmas with parameters in this range that are based on only a small number of sessions.

2 Preliminaries

In the experiments we analyze, participants played a sequence of repeated prisoner's dilemma games with perfect monitoring.[1] The game parameters were held fixed within each session, so each participant only played one version of the repeated game. The treatments all had randomly chosen partners and a random stopping time, so the discount factor δ corresponds to the probability that the current repeated game continues at the end of the current round. (We will refer to the "rounds" of a given repeated game, and call each repeated game a new "supergame.")

We represent the prisoner's dilemma with the following strategic form, where $g, l > 0$ and $g < l + 1$. Here g measures the gain to defection when one's opponent cooperates, l measures the gain to defection when one's opponent defects, and $g < l + 1$ implies that the efficient outcome is (C, C).

[1] There are many more experiments on this case than on the prisoner's dilemma with implementation errors or imperfect monitoring.

	C	D
C	$1, 1$	$-l, 1+g$
D	$1+g, -l$	$0, 0$

Figure 1: The Prisoner's Dilemma

Standard arguments show that "Cooperate every round" is the outcome of a subgame-perfect equilibrium if and only if it is a subgame-perfect equilibrium (SPE) for both players to use the strategy "Grim": Play C in the first round and then play C iff no one has ever played D in the past. This profile is a SPE iff

$$1 \geq (1-\delta)(1+g) \iff \delta \geq g/(1+g) \equiv \delta^{\text{SPE}}.$$

Note that the loss l incurred to (C, D) does not enter in to this equation, because the incentive constraints for equilibrium assume that each player is certain their opponent uses their conjectured equilibrium strategy.

Applications of repeated games often assume that players will cooperate whenever cooperation can be supported by an equilibrium,[2] but this hypothesis has little experimental support. Instead, the level of cooperation in repeated game experiments can be better predicted by measures that reflect uncertainty about the opponents' play. In particular, our work uses the composite parameter

$$\Delta^{RD} \equiv \delta - \delta^{\text{RD}} = \delta - (g+l)/(1+g+l)$$

that was suggested by Blonski, Ockenfels and Spagnolo (2011).[3]

Inspired by previous work and descriptive evidence we present later, we develop a very simple model that assumes all individuals use memory-1 strategies, and moreover that these strategies differ across treatments and supergames only with respect to play in the initial round of each supergame. We assume that the probability of cooperation in the initial round of each supergame s, $p_i^{initial}(s)$, depends on the game parameters and the effect of individual experience in previous supergames, $e_i(s)$, according to

$$p_i^{initial}(s) = \frac{1}{1+\exp\left(-(\alpha + \beta \cdot \Delta^{RD} + e_i(s))\right)}. \tag{1}$$

[2]See e.g. Rotemberg and Saloner (1986), Athey and Bagwell (2001), and Harrington (2017).

[3]Rand and Nowak (2013) regresses first-round cooperation on Δ^{RD}; Dal Bó and Fréchette (2011, 2018) instead uses the size of the basin of attraction for AllD. We tried using that instead of Δ^{RD}, but the results were slightly worse.

Thus initial-round behavior in the model is driven by two components: a direct effect of game parameters, captured by the linear function $\alpha + \beta \cdot \Delta^{RD}$, and the effect of reinforcement learning, captured by individual experience $e_i(s)$.[4]

To model learning, we suppose that after each supergame $s \in \{1, 2, ...\}$, $e_i(s)$ is updated according to

$$e_i(s) = \lambda \cdot a_i(s-1) \cdot V_i(s-1) + e_i(s-1), \tag{2}$$

where $a_i(s) = 1$ if player i's initial round action in supergame s was C and $a_i(s) = -1$ if player i's initial round action was D, $V_i(s)$ is the total payoff received in that supergame, λ determines the strength of the learning, and $e_i(1) = 0$.[5] Thus, reinforcement of cooperation or defection in the initial round depends on the resulting supergame payoffs, while the direct influence of Δ^{RD} is constant across supergames.

We assume that behavior at non-initial rounds follows a memory-1 mixed strategy that is constant across individuals, treatments, and time. Let $h \in \{CC, DC, CD, DD\}$ denote a memory-1 history, and let σ_h be the probability of cooperation at one of these histories. Following Breitmoser (2015), we assume these correspond to a "semi-grim" strategy, i.e. that $\sigma_{CC} > \sigma_{DC} = \sigma_{CD} > \sigma_{DD}$. (In section 5.3 we relax this assumption and consider multiple extensions, but this does not improve predictions.) The interpretation of this mixed strategy is not necessarily that participants are consciously randomizing their behavior with these probabilities, and would say so if asked. Instead, they might be unsure what to do as they play the game.

In total, our model has 6 parameters, $(\alpha, \beta, \lambda, \sigma_{CC}, \sigma_{CD/DC}, \sigma_{DD})$. We call this "IRL-SG" for "the initial-round learning with semi-grim strategies model."

Importantly, to predict average cooperation, we use simulations that suppose all participants use learning rules of the form (1) and (2) to determine the evolution of each simulated participant's experience and play. We do *not* use data on the payoffs received by the actual participants in the experiments.

[4]The additive form of reinforcement is standard in the literature, see e.g. Erev and Roth (1998).

[5]The alternative specification where learning responds to the average payoff in a supergame instead of the total does worse in-sample than our model's cross-validated error, so we did not try to evaluate its outsample performance. This suggests that learning between supergames is stronger when the supergames are longer.

3 Prior Work

Blonski, Ockenfels and Spagnolo (2011) assumes that play will always correspond to an equilibrium of the repeated game, and that there is a functional form that determines which equilibria are possible for each parameter configuration. It then uses five axioms on this functional form[6] to conclude all players will defect if Δ^{RD} is negative. Blonski and Spagnolo (2015) shows that the sign of Δ^{RD} corresponds to whether Grim is risk dominant in a 2x2 matrix game with the strategies Grim and Always Defect, and also refines the prediction that players will cooperate to a prediction of which cooperative strategies the players will use. Chassang (2010) shows that the risk-dominance threshold characterizes the limits of equilibria of incomplete-information repeated games with an exit option. Because players know each other's equilibrium strategies, and have the option of leaving a match early, the setting is rather different than the one we study here.

On the empirical side, Blonski, Ockenfels and Spagnolo (2011), Rand and Nowak (2013), and Blonski and Spagnolo (2015) show that the average cooperation rates in a session are increasing in Δ^{RD}. Dal Bó and Fréchette (2018) shows that the sign of Δ^{RD} is much more correlated with high cooperation rates than is the sign of $(\delta - \delta^{\text{SPE}})$.[7] These papers did not treat cooperation as a prediction problem, and did not propose a mechanism to explain the observed dependence on Δ^{RD}.

Several papers estimate the strategies used by participants treatment by treatment on the assumption that each participant uses a fixed strategy either in the entire session or in the latter part of it. The papers that assume participants use pure strategies consistently find that most of the behavior can be captured by the strategies AllD (Always Defect), TFT (Play C in the initial round of a supergame, and thereafter play the action your partner played in the previous round), Grim (Play C in the initial round and thereafter play D if either partner has ever defected), and for lower values of Δ^{RD}, D-TFT (play D in the initial round and thereafter play what your partner played in the previous round).[8]

[6]The axioms include an additive separability condition and the requirement that g and l have equal weight.

[7]Dal Bó and Fréchette (2011) find that cooperation is correlated with a different function of g, l and δ, namely $\frac{(1-\delta)l}{1-(1-\delta)(1+g-l)}$. This is highly correlated with Δ^{RD}, but when we used it instead we obtained less accurate predictions.

[8]See for example Dal Bó and Fréchette (2011); Fudenberg, Rand and Dreber (2012); Dal Bó and Fréchette (2018). Fudenberg, Rand and Dreber (2012) shows that longer memories are used when

More recent within-session studies find evidence for the use of "memory-1" mixed strategies that depend only on the actions played in the previous round. Breitmoser (2015) finds that semi-grim strategies fit play after the initial round better than pure strategies do. Backhaus and Breitmoser (2020) argues that a combination of AllD and two types of semi-grim with the same non-initial play best fits behavior when the mixtures of the types as well as their play is estimated treatment by treatment. In Online Appendix E, we adapt our model to predict the next action played given the history so far and find that it performs well.

Dal Bó (2005) and subsequent work shows that behavior changes between the first and last supergame in a session. Moreover, Dal Bó and Fréchette (2011) argues that δ has no apparent effect on behavior in the first supergame, but a substantial impact on later supergames. Similarly, the difference between treatments increases over time, with average cooperation going down in games where no cooperative SPE exist, and going up in games where Δ^{RD} is high. A common explanation for the observed time trends is that participants learn from feedback over the course of a session, and choose their supergame strategies based on outcomes in the previous supergames. [9]

The literature also establishes two other empirical regularities that are related to learning. First, cooperation increases when the realized supergame lengths are longer than expected, and decreases when they are shorter than expected. Engle-Warnick and Slonim (2006) and Dal Bó and Fréchette (2011, 2018) find that cooperation in the initial round increases if the previous supergame was longer than expected, and Mengel, Weidenholzer and Orlandi (2021) finds that this effect is persistent: cooperation later in a session is higher when the early supergames are longer than expected. Second, Dal Bó and Fréchette (2011, 2018) find that initial-round cooperation is higher if the player's partner cooperated in the first round of the previous supergame. These two effects point to a model where some form of learning or reinforcement drives play in the initial rounds.

---

the intended actions are implemented with noise and only the realized actions are observed. Other papers elicit supergame strategies directly, e.g. Romero and Rosokha (2018), Dal Bó and Fréchette (2019), and Romero and Rosokha (2019) and argue that the elicited strategies are a good description of how people play in more natural settings. None of these papers allow for learning or produce out of sample predictions.

[9] Dal Bó and Fréchette (2011) considers a simple belief learning model involving only TFT and AllD, Mengel, Weidenholzer and Orlandi (2021) consider a similar learning model that also incorporates learning about expected supergame length, and Erev and Roth (2001), Hanaki et al. (2005) and Ioannou and Romero (2014) study learning *within* supergames.

The larger literature on learning in game theory experiments has been focused on one-shot games, for example in (Cheung and Friedman, 1997; Erev and Roth, 1998; Camerer and Ho, 1999), and has not emphasized the issue of out-of-sample prediction. Fudenberg and Liang (2019) and Wright and Leyton-Brown (2017) study ways to predict initial play in matrix games, but don't consider learning.

4 Summary of the data

We analyze the data from the meta-analysis in Dal Bó and Fréchette (2018), which included papers on repeated prisoner's dilemma experiments with perfect monitoring, deterministic payoffs, and constant parameters within a session that were published by 2015. We use only the treatments with $\delta > 0$, and augment this data with data from sessions that match these criteria from four papers published since then (Aoyagi, Bhaskar and Fréchette (2019); Dal Bó and Fréchette (2019); Proto, Rustichini and Sofianos (2019); Honhon and Hyndman (2020)). This increases the number of sessions by approximately 60%. Our resulting data set contains observations from 17 papers, 28 different treatments,[10] and 161 incentivized experimental laboratory sessions, containing 2,612 distinct participants and 232,298 individual choices. Here we highlight some aspects of the data that are of particular relevance to our work.

The discount factors ranged from 0.125 to 0.95, with almost all at least 0.5. In 20 of the sessions, $\delta < \delta^{\text{SPE}}$, so no cooperation can occur in a subgame perfect equilibrium. In 28 sessions, cooperation can be supported by a SPE, i.e. $\delta > \delta^{\text{SPE}}$, but $\delta < \delta^{\text{RD}}$, so cooperation is not risk dominant in the sense of Blonski, Ockenfels and Spagnolo (2011). In the remaining 113 sessions, $\delta > \delta^{\text{RD}}$ or equivalently $\Delta^{RD} > 0$.

The average rate of cooperation over all sessions was 44.1%. It was 10.5% for games where $\delta < \delta^{\text{SPE}}$, 18.6% for $\delta^{\text{SPE}} < \delta < \delta^{\text{RD}}$, and 53.6% for $\delta > \delta^{\text{RD}}$.

Table 1 shows average play after the different memory-1 histories, and the histories' frequencies. We see that the CD and DC histories are only a small subset of observations, roughly 14% combined. Furthermore, we see that the average behavior is close to the semi-grim memory-1 mixed strategy from Breitmoser (2015), with the difference that the probability of cooperation is slightly higher after DC than after

[10]Following Dal Bó and Fréchette (2018), we consider experiments in different labs to be the same treatment if they had the same normalized parameters. The total number of unique paper and parameter combinations is 47.

CD.

Table 1: Average cooperation rate after different memory-1 histories.

History	Avg C	N
CC	96.6%	59,435
CD	30.6%	16,706
DC	33.2%	16,706
DD	5.2%	74,621
initial	47.1%	64,830
Total	44.1%	232,298

As shown in Online Appendix A, the average rate of cooperation in the initial round of the first supergame is increasing in Δ^{RD}. Moreover, the way initial-round cooperation changes over the course of a session varies with Δ^{RD}: For $\Delta^{RD} > 0.15$, initial-round cooperation rates increase over the course of a session; for $\Delta^{RD} < 0$ they decrease, and in sessions where $0 < \Delta^{RD} < 0.15$, they remain roughly constant at around 50%. This pattern is much less sharp after the other memory-1 histories. This suggests that the differences in average cooperation across treatments is primarily driven by differences in the behavior at the initial round. Moreover, we can predict cooperation in a given match relatively well if we know the outcome of the initial round. Specifically, for each participant and supergame that lasted at least two rounds, we let the outcome variable be the average cooperation by that participant in the non-initial rounds. As reported in table 1 of the Online Appendix, we then consider three different regressions. In the first we only condition on the outcome of the initial round, in the second we add the game parameters (g, l, δ, and Δ^{RD}), and in the last we remove the outcome of the initial round. The difference in R^2 between the first and second regression is less than 0.01, while the third regression has a much lower R^2. Doing the same regressions but with second-round average cooperation as the outcome does not change the picture.

5 Predicting Cooperation

Our goal in this paper is to find a model that captures the essence of behavior across treatments, not to explain everything that goes on in a single experimental session. To that end, we try to develop models that can successfully predict cooperation levels

in a repeated game experiment before the experiment is run, based on the game parameters, the number of repeated games played, and their lengths, without using any data from the experiment itself. We evaluate models based on their estimated out-of-sample predictive performance as measured by cross-validated mean squared error (MSE). Using cross-validation helps prevent overfitting, and makes sure that the regularities we find actually improve predictions. In general, out-of-sample predictions will favor models that rely on stable predictors and do not overfit the data, and such models are typically simpler than those that give the best in-sample fit.

In addition, using out-of-sample prediction error as the benchmark makes it easy to compare models of different complexities, because the out-of-sample prediction criterion endogenously penalizes models that are too complex. We also report the relative improvement of the models compared to a constant prediction benchmark, in order to get a better sense of how big the differences are.[11]

5.1 Predicting Cooperation with Learning

To make predictions with the IRL-SG, and with more general variants of it that we discuss later, we simulate populations playing the different sessions assuming they behave as the model specifies. In particular, we use the simulations to generate the experience levels e_i, and do not use data on the actual initial round actions or the payoffs that people actually received. We initialize a large population of individuals, all with $e_i(1) = 0$. For a given specification of parameters of the learning model, we randomly match these simulated individuals to play a sequence of supergames. After each supergame, the individual experiences are updated according to equation (2), using the simulated values. The simulated individuals are then randomly re-matched and play the second supergame for the number of rounds it was played in the experimental session. So it continues until we have simulated a population playing exactly the same sequence of supergames as in the experimental session, updating the $e_i(s)$ after each supergame.

Once we have simulated a population, we can calculate either average cooperation or the time path of cooperation, that is the percentage of participants who cooperate in each round $1, 2, ...$ of any supergame in a given treatment. We use the simulations

[11] This use of a simple prediction rule as a benchmark is inspired by the completeness measure of Fudenberg et al. (2021), but we do not have enough data to make a good estimate of the problem's irreducible error as the completeness measure requires.

as predictions, and compute the approximate prediction losses and associated standard errors.

We estimate the learning models based on the time path of cooperation, even when predicting average cooperation. That is, we find the parameters that best predict the time path of cooperation in the training set, and use those parameters to predict both the average cooperation and the time path of cooperation in the test sets. This way, we use more of the data to estimate the model.

Appendix A gives a detailed description of the numerical process. In Online Appendix C we evaluate this estimation procedure on data simulated under different assumptions about how people actually behave, and confirm that it should work on a data set of the size we have.

Our main learning model, presented earlier in equations (1) and (2), assumes all agents use the same learning rule, which is an oversimplification. However, agents with the same learning rule can behave differently as the result of different experiences, and as we will see allowing heterogeneous learning rules does not improve our predictions.

The restriction to memory-1 strategies is motivated by past work and also by the machine learning analysis in Online Appendix E. The assumption that play across treatments is the same except in the initial rounds is motivated by the descriptive statistics. We relax the assumption of fixed behavior at non-initial histories in section 5.3, which considers a more richly parameterized model that lets play at these histories depend on Δ^{RD}. We also consider variations of the IRL-SG that generalize the learning part of the model in various ways, and alternative models where individuals learn which pure strategy to use. None of these extensions improve predictions.

5.2 Results

We now compare the performance of IRL-SG to that of OLS, Lasso, SVR (support vector regression), and GBT (gradient boosting trees). When we use machine learning to predict the average cooperation level in a session, each session is a single data point. The feature set consists of Δ^{RD}, the game parameters (g, l, δ), the total number of rounds played in the session, the number of supergames played in the session, an indicator variable for whether $\Delta^{RD} > 0$, the difference between expected and realized supergame lengths in the first third of the session[12], the total difference between

[12]Mengel, Weidenholzer and Orlandi (2021) argues that the realized length in the first third of the supergames has an oversized impact on later cooperation.

expected and realized supergame lengths, and some interaction terms.[13]

When we predict the time path of cooperation, a data point is the average (across participants) cooperation level of each round of each supergame in a session; this gives a total of 15,598 data points, though the data within each session is highly correlated. For the feature set of the time path, we add an indicator for the initial round, the round number, and the supergame number, along with some interaction terms. We replace the features about realized supergame lengths with the cumulative difference in expected and realized supergame lengths, and the difference between expected and realized length of the previous supergame.

Table 2 shows the out-of-sample prediction errors for average cooperation for the IRL-SG as well as for SVR (the best performing atheoretical predictor here) and OLS. We see that our learning model performs much better than simple OLS on Δ^{RD}, and in fact slightly better than the harder-to-interpret machine learning algorithms. This is in part due to the fact that our data set is relatively small by machine learning standards. Also, the IRL-SG is better able to incorporate the effect of the realized supergame lengths, as can be seen by the fact that when we redo the estimations without using the realized supergame lengths as data, the out-of-sample MSE for the IRL-SG and the best ML method both increase to 0.0158.

Model	Avg C MSE	S.E.	Relative Improvement
Constant	0.0517	(0.0040)	-
OLS on Δ^{RD}	0.0189	(0.0020)	63.4%
SVR	0.0145	(0.0016)	71.9%
IRL-SG	0.0138	(0.0015)	73.3%

Table 2: Out-of-sample prediction MSE for average cooperation

To estimate the out-of-sample prediction errors, we use 10-fold cross validation: We divide the sessions into 10 different folds, with data split on the level of the session, so each observation is predicted using only data from other sessions. For each fold, we use the other nine folds as a training set to estimate the parameters, and make predictions on the test fold using those parameters.[14] To estimate the standard errors of the estimated mean squared error (MSE), we do 10 different such

[13]The alternative method of training the ML algorithms on time paths, and using time path predictions to predict average cooperation, yields similar but slightly worse results.

[14]See, e.g., Hastie, Tibshirani and Friedman (2009) for an explanation of cross validation.

cross validations, leading to 10 predictions of each session, to capture the randomness from the assignment to different folds. We then bootstrap to get standard errors. By using the same folds for all models we can perform pairwise comparisons. Pairwise tests are presented in Appendix D. According to those pairwise tests, our learning model is indeed significantly better than an OLS on Δ^{RD}, though its improvement on SVR is too small to be significant.

In table 3 we see similar results for predicting the time path of cooperation. Not only is IRL-SG better than the alternatives at predicting average cooperation in a session, it is also better at predicting the time path of cooperation:

Model	Time-path MSE	S.E.	Relative Improvement
Constant	0.0775	(0.0050)	-
OLS on Δ^{RD}	0.0398	(0.0025)	48.6%
GBT	0.0321	(0.0020)	58.6%
IRL-SG	0.0309	(0.0020)	60.1%

Table 3: Out-of-sample prediction loss for predicting the time path of cooperation.

Figure 2 shows the out-of-sample predictions and actual values of cooperation in the initial round of the first 20 supergames. (We plot the initial round to reduce the noise introduced by changing supergame lengths.) To get the out-of-sample predictions for the figure, we use a single 10-fold cross-validation split and then predict each session's time path with the parameters estimated without data from that session. The learning model predicts the general pattern well, but it slightly underestimates the level of cooperation for intermediate values of Δ^{RD}.

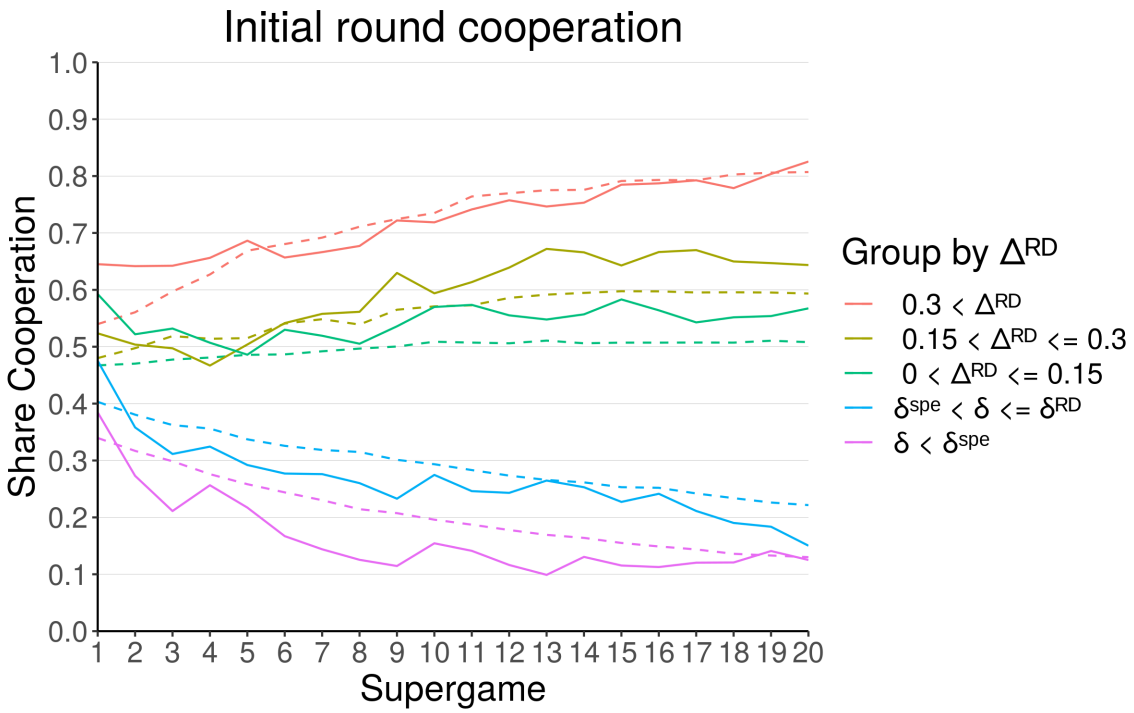

Figure 2: Actual (solid line) and out-of-sample predicted (dashed line) initial-round cooperation by supergame for sessions of at least 20 supergames .

The next table shows the average of the estimated parameters; the standard deviations show how much theses parameter estimates vary between folds.

Parameter	α	β	λ	p_{CC}	$p_{CD/DC}$	p_{DD}
Average	-0.268	1.291	0.182	0.995	0.355	0.012
Standard Deviation	(0.061)	(0.160)	(0.036)	(0.002)	(0.026)	(0.006)

Table 4: Parameter estimates

From here on, when we analyze the behavior of the model and discuss parameter values, we will be using these average estimates.

To interpret the parameter estimates, recall that experience is updated according to

$$e_i(s) = \lambda \cdot a_i(s-1) \cdot V_i(s-1) + e_i(s-1).$$

which then enters into the probability of initial round cooperation via

$$p_i^{initial}(s) = \frac{1}{1 + \exp\left(-(\alpha + \beta \cdot \Delta^{RD} + e_i(s))\right)}.$$

The estimated $\alpha = -0.268$ means that for $\Delta^{RD} = 0$, about 43.3% of participants would cooperate in the first round of their first supergame. With $\Delta^{RD} = 0.1$, the probability of cooperation in the first round of the first supergame increases, but only to 46.5%, and even when Δ^{RD} increases to 0.3 first-round cooperation only increases to 53%.

In contrast, $\lambda = 0.182$ implies a strong learning effect. As an example, consider the case where $g = l = 2$ and $\delta = 0.8$, so $\Delta^{RD} = 0$. If the first supergame an individual i plays goes the expected 5 rounds and both partners cooperate all 5 rounds, then i's probability of first-round cooperation goes from 43.3% to 65.6%. An individual j experiencing DC in the first round and DD in the remaining 4 rounds gets a payoff of 3, which implies that their initial-round cooperation would decrease to 30.6%.

To get a sense of the relative importance the model assigns to Δ^{RD} and learning, we compute the Shapley values of these terms in a decomposition of the variance of predicted initial play in the last supergame. Since Δ^{RD} and $e_i(s)$ are correlated, and enter into the model in a non-linear fashion, it is non-trivial to do this decomposition; our use of the Shapley value follows Lundberg and Lee (2017) and Molnar (2019). As we show in Appendix C, this decomposition suggests that in the last supergame of each session, approximately 88% of the variation between treatments is driven by learning as opposed to the direct influence of the game parameters.

5.3 Comparison with Alternative Models

We considered variants of our model that add more parameters, and also models where participants learn to play pure strategies. We present the most relevant comparisons here, and include further models and complete results in Appendix B. The variant with two types of agents with different parameters does very slightly better than the IRL-SG in terms of cross-validated prediction error (but not significantly so); all of the other alternatives do worse and in particular all models based on pure strategy learning are significantly worse.

Learning at all memory-1 histories. So far, we have restricted learning to the initial round, and kept behavior at non-initial rounds constant, both across time and treatments. To relax this and allow learning at all memory-1 histories, we track experience $e_i(h, t)$ at each memory-1 history h, where t is now a time variable running over all rounds and all supergames.

The intuition behind this model is similar to the intuition for the IRL-SG. If a player choose action a at memory-1 history h, she is more likely to choose a the next time she observes h if it lead to a good outcome, and less likely if it lead to bad outcome. To define good or bad outcomes, we let $V_i(t)$ be the total payoff from time t to the end of the corresponding supergame. If a player cooperated after a CD history in round 3 in a supergame, and both players ended up cooperating in the remaining two rounds of that supergame, the corresponding $V_i(t)$ is 2. This is independent of the payoffs in earlier rounds in that supergame. Furthermore, since $V_i(t)$ is not known before the supergame has ended, we only consider learning across supergames.

In other words, the experience at memory-1 history h is updated when h occurs using the rest of the supergame payoff $V_i(t)$ according to

$$e_i(h, t+1) = \begin{cases} \lambda \cdot a_i(t) \cdot V_i(t) + e_i(h,t) & \text{if } h(t) = h \\ e_i(h,t) & \text{if } h(t) \neq h \end{cases}$$

where $h(t)$ is the memory-1 history at time t.

The probabilities of cooperation are updated at the beginning of each supergame, and remain constant in its subsequent rounds. So the probability to cooperate at memory-1 history h is given by

$$p_i(h,t) = \begin{cases} \frac{1}{1+\exp(-(\alpha^h + \beta^h \cdot \Delta^{RD} + e_i(h,t))} & \text{if } r(t) = 1 \\ p_i(h, t-1) & \text{if } r(t) > 1 \end{cases}$$

where $r(t)$ denotes the round at time t. This model thus has 11 parameters. A last variation of this model allows for two different learning rates: Learning in the initial round happens with $\lambda_{initial}$, and learning for the memory-1 histories is reinforced with λ_1, which increases the number of parameters to 12.

The IRL-SG with two types. We assume that there are two different types with different parameters, and one parameter that determines the shares of the two types in the population, which adds seven parameters. [15]

Pure strategy belief learning. The pure strategy belief learning model in Dal Bó and Fréchette (2011) assumes that all participants follow either TFT or

[15] There is not much evidence of heterogeneity in our data, but other works shows that some forms of heterogeneity do matter; for example Proto, Rustichini and Sofianos (2019) shows that groups of more intelligent people are more likely to cooperate. This suggests we might be able to improve predictions if we had demographic information about the participants in each session.

AllD. Each participant has beliefs about how common TFT and AllD are in the population, which the participant updates based (only) on the opponents' moves in the initial rounds, and uses them to calculate the expected values from playing TFT or AllD. Given these expected values, the participant's choice of whether to play TFT or AllD in the following supergame is given by a logistic best reply function. We extend this model to allow for across-treatment prediction, increasing the original 6 parameters to 8. We also consider a version of the pure strategy model with symmetric implementation errors. A more complete description of the model can be found in Online Appendix B

Pure strategy reinforcement learning. In the pure strategy reinforcement learning model, we consider reinforcement learning over the pure strategies AllD, Grim, and TFT. Each of the pure strategies k starts with an initial attraction $A_k(1)$. At the beginning of each supergame s, the individual samples the pure strategy to use according to

$$p_k(s) = \frac{\exp\left(\lambda A_k(s)\right)}{\sum_{l \in \{TFT, AllD, Grim\}} \exp\left(\lambda A_l(s)\right)}$$

where $p_k(s)$ is the probability of using pure strategy k in supergame s, and λ denotes the sensitivity. Let $k(s)$ denote the pure strategy used in supergame s, then after supergame s, the attraction of the pure strategy used is updated according to

$$A_k(s+1) = \begin{cases} A_k(s) + V(s) & \text{if } k(s) = k \\ A_k(s) & \text{otherwise.} \end{cases}$$

The initial attractions are given by linear functions of Δ^{RD}, i.e., $A_k(1) = \alpha_k + \beta_k \Delta^{RD}$.

We then extend the model to allow for symmetric errors or "trembles" ε when implementing a pure strategy. In other words, if an individual is following TFT and the previous history is DC, the probability of cooperating is $1 - \varepsilon$, instead of 1, and similarly for other pure strategies and histories. In total this model has 7 (8) parameters without (with) symmetric trembles.

Results In table 5 we see a comparison between some of the alternatives considered in this subsection. Neither pure strategy model does as well as IRL-SG. Appendix B gives a complete table of these comparisons, and the results for the time-path

prediction problem, which show a similar relationship between the models. Pairwise tests for significance are presented in Appendix D.

Model	Avg C MSE	S.E.	Relative Improvement
Pure strategy belief learning w/ trembles	0.0191	(0.0020)	63.0%
Pure strategy reinf. learning w/ trembles	0.0175	(0.0020)	66.1%
Learning at all memory-1	0.0139	(0.0016)	73.1%
IRL-SG	0.0138	(0.0015)	73.3%
IRL-SG, two types	0.0137	(0.0015)	73.5%

Table 5: Out-of-sample prediction loss (MSE) of average cooperation

Taken together, this suggests that the simple IRL-SG captures most of the predictable regularity in cooperation rates. Introducing heterogeneity improves predictions at most marginally, as does learning or flexibility at non-initial rounds. If fact, extending our model often seems to lead to slightly worse out-of-sample performance, most likely due to overfitting. We also see that assuming pure strategy learning models does not lead to good predictions.

To further test how well IRL-SG captures the predictable aspects of the data, in Online Appendix F.1 we consider different combinations of IRL-SG predictions and ML-methods. In the first approach, we use the predictions of the IRL-SG as a feature for the ML-methods. In the second, predictions of initial-round behavior are taken from the IRL-SG. We then use machine learning to predict play in subsequent rounds. Neither approach improved predictions by a non-trivial amount.

In Online Appendix E we instead consider the problem of predicting the next actions taken by an individual player. Here we see some advantage to allowing different types. However, that advantage is much smaller for our learning model than for models without learning, and a single IRL-SG performs as well as a mixture model that estimates the shares of 11 different pure strategies on each treatment.

5.4 Understanding the Model

Our simple learning model is able to accurately predict average cooperation and the time path of cooperation, while holding fixed the strategies used in the non-initial rounds. We now briefly discuss the roles of the more important parts of our model.

Reinforcement of Initial Actions. For each session, let $\pi(C)$ be the average supergame payoff received by participants who cooperated in the initial round and define $\pi(D)$ analogously. Figure 3 demonstrates the correlation between $\pi(C) - \pi(D)$ and Δ^{RD} in the experimental and simulated data sets. The simulated data is generated with 100 populations of size 16 for each session, simulated on sessions with the same supergame lengths as in the data. The populations size 16 is chosen based on the average population size in the data.

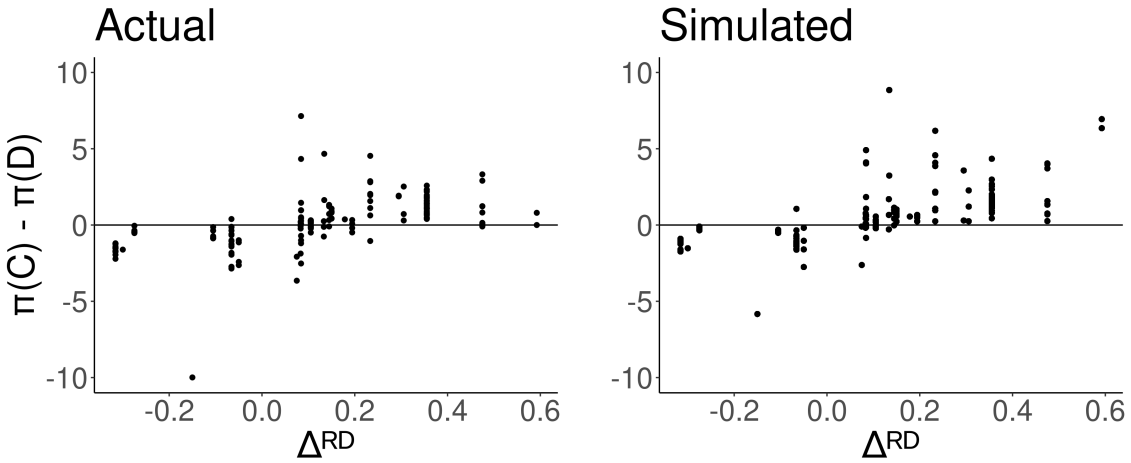

Figure 3: Average empirical (left) and simulated (right) difference between total payoff in supergames where the participant cooperated and defected. Each dot corresponds to one session.

For $\Delta^{RD} < 0$, defection is reinforced more strongly than cooperation in all but one session. For positive but low values of Δ^{RD}, the difference in reinforcement $\pi(C) - \pi(D)$ is centered around 0, so cooperating and defecting are on average equally reinforced. This helps explain why we see no clear time trends in the sessions where $0 < \Delta^{RD} < 0.15$.

Even though $\pi(C) - \pi(D)$ is correlated with Δ^{RD}, our learning model makes better predictions than using Δ^{RD} directly. This suggests that the dynamics of our learning model help capture some additional forces that determine cooperation, such as how many supergames were played and their realized lengths. Indeed, as we observed above, our model does make better use of the realized supergame lengths than our ML algorithms do. In particular, our model succeeds in replicating the empirical fact that there is more cooperation when the realized supergames are longer. Intuitively, this is because regardless of the distribution of opponent play, the potential reward

from initially playing C is increasing in the realized supergame length, as it comes from triggering many future rounds of where both play C, while the reward from an initial D is obtained immediately.

The composition of Δ^{RD}. To see how well Δ^{RD} captures the effects of the individual game parameters, we simulate populations with the same Δ^{RD} but with different game parameters. Looking at the formula for Δ^{RD}, we see that it is constant as we vary g and l provided we keep $g + l$ fixed. The next figure plots predicted average cooperation for $g+l = 3$, with δ chosen so that $\Delta^{RD} \in \{0, 0.1, 0.2\}$. ($\delta = 0.75$ gives $\Delta^{RD} = 0$.) For each of the treatments we consider (a dot in the figure), we simulate 1000 populations of 16 playing 27 supergames (the average in the data). As we see in Figure 4, our model predicts that the values of g and l have an effect that is not captured by the composite parameter Δ^{RD}. In particular, increasing $g - l$ for fixed $g + l$ seems to dampen the effect of Δ^{RD}. In other words, higher $g - l$ results in decreased cooperation if Δ^{RD} is high, but increased cooperation if Δ^{RD} is low.[16]

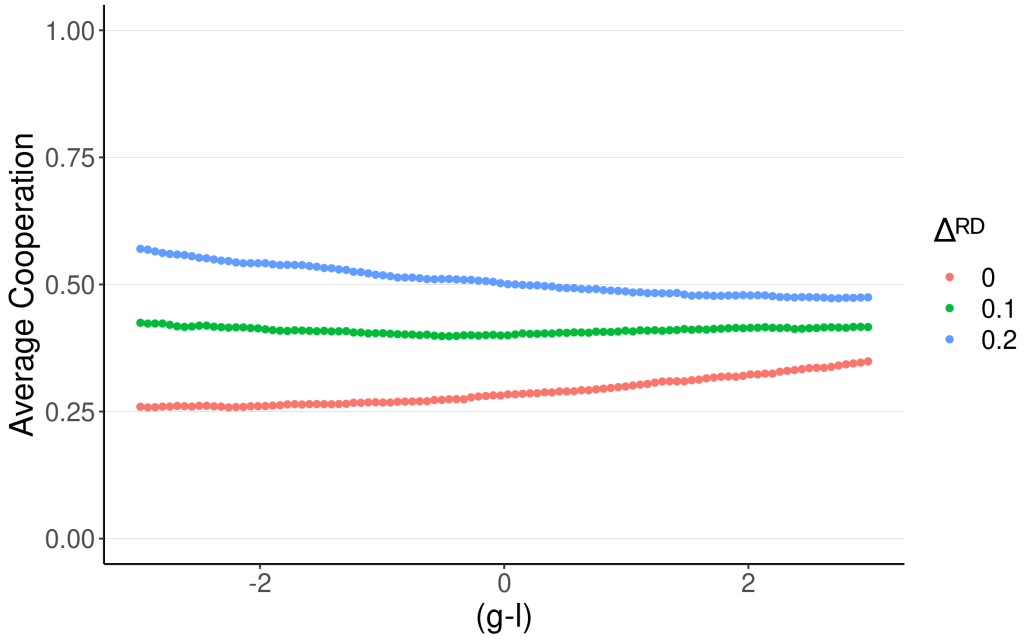

Figure 4: Predicted average cooperation over 27 supergames for fixed $g+l$ but varying $g - l$.

[16]There are not enough experiments with the same Δ^{RD} and different g, l to directly test this prediction:We have 2 treatments for $\Delta^{RD} = -.05$, with 3 sessions of one parameter constellation and 1 of the other, and 4 treatments for $\Delta^{RD} = .0833$, but only 5 sessions in total.

Between session variation and Δ^{RD}. In the IRL-SG, behavior is reinforced based on experience. Since both behavior and realized supergame lengths are random, there is noise in the realized experiences. When $\delta < \delta^{SPE}$, there is a unique equilibrium and over time experiences should drive all individuals to defection. When $\delta \geq \delta^{SPE}$, randomness could drive the population to either cooperation or defection. Thus for low values of Δ^{RD} we predict and see consistently low cooperation rates with relatively small between session variance. For intermediate values, we should see very high levels of between session variance, since small random differences can have big effects. Lastly, for high values of Δ^{RD} we should see more variance than for very low values of Δ^{RD}, but less than for intermediate values of Δ^{RD}.

In the Table 6 we see that this pattern is observed both in the experimental and in the simulated data. For the simulated data, 100 sessions with population size 16 were simulated for each session in the experimental data, each with the same realized sequence of supergame lengths. For most values of Δ^{RD}, we only have a few sessions. We therefore group the sessions into similar values of Δ^{RD}. For each of the group we report the 5%-quantile, the median, and the 95%-quantile.

	Actual			Simulated		
Δ^{RD}-group	Q05	Median	Q95	Q05	Median	Q95
$\delta < \delta^{SPE}$	0.03	0.14	0.26	0.08	0.13	0.29
$\delta^{SPE} < \delta \leq \delta^{RD}$	0.03	0.18	0.30	0.09	0.19	0.30
$0 < \Delta^{RD} \leq 0.15$	0.15	0.40	0.70	0.20	0.37	0.66
$0.15 < \Delta^{RD} \leq 0.3$	0.30	0.55	0.74	0.25	0.49	0.67
$0.3 < \Delta^{RD}$	0.44	0.65	0.83	0.53	0.66	0.85

Table 6: Across session variation in average cooperation for different groups of Δ^{RD}.

This large variation in outcomes for intermediate values of Δ^{RD} suggests that caution should be taken when interpreting experimental studies of the Prisoner's Dilemma. Most experiments have few sessions per treatment condition, and learning effects can give rise to large variations between sessions even when the individuals are ex-ante identical.

Non-initial play. The performance of the IRL-SG suggests that learning does not lead agents to adjust their play very much at non-initial rounds. Here is one conjecture about why this might be: Suppose there is a cognitive cost associated with learning from experience and adjustment to game parameters. In that case, we

should expect learning and adjustment to happen where the relative payoff is the greatest, which is in the initial round.[17] To test this intuition, we assume that all other individuals behave according to our estimated IRL-SG, and then calculate the potential gain from learning and adjustment at different histories. We first consider a baseline fixed memory-1 model without any learning or adjustment to Δ^{RD}. We then consider extensions where learning and adjustment are added to initial and/or non-initial memory-1 histories. The best fixed model achieves a normalized average payoff per session of 43.6; the best IRL-SG achieves 54.3 and the model with learning at all histories and two learning rates only increases that to 55.4.

6 Extrapolating to Longer Experiments

Due to practical constraints, experiments on the PD are of limited duration, but as researchers we are also interested in what would happen over a longer run. Our learning model lets us make predictions of what would happen in experiments with a a longer time horizon than those in our data set. This gives us some sense of which properties of observed lab play may carry over to repeated games that are played more than is practical in the lab.[18]

6.1 Extrapolating within observed sessions

Before exploring the implications of the learning model for long-run play, we want to test how well it can extrapolate to longer sessions than it is trained on. To do this, we use the same cross-validation folds as earlier, so that data from a given session is either in a training fold or a test fold but not both. We then use the first halves of the training sessions to estimate the parameters, and use the estimated model to predict the second half of the sessions in each test set. This is a way of approximating how accurate our predictions would be for experiments that are twice as long as the ones in the sample.

We estimate the parameters of the different models in the first half of the session,

[17] There are more initial histories than DC and CD histories, and the evaluation of non-initial histories may depend on outcomes in prior rounds. Also, given the path dependence of play, our prior is that the initial decision is the most important.

[18] The median number of supergames played in our data is 21 and in 90% of sessions participants play between 6 and 65 supergames.

and use them to predict average cooperation in the second half. To do this, we first predict the time path and calculate the resulting average cooperation. We do this for both the IRL-SG and ML algorithms. Table 7 displays the cross-validated MSE.

Model	Avg C MSE	Avg C S.E.
Constant	0.0695	(0.0055)
SVR	0.0285	(0.0030)
Lasso	0.0284	(0.0030)
OLS on Δ^{RD}	0.0281	(0.0032)
GBT	0.0266	(0.0027)
IRL-SG	0.0220	(0.0026)

Table 7: Prediction loss (MSE) estimating on 1st half and evaluating on 2nd half.

The table shows that the learning model is better at extrapolating to longer supergames than our atheoretical black-box algorithms or simple OLS. Appendix D confirms that this difference is significant using pairwise tests. This might be due to our particular ML implementations, but it is also true that atheoretical prediction algorithms can have trouble extrapolating to slightly different settings. A more structured model that encodes some intuition or knowledge about the problem domain can sometimes better extrapolate to related prediction problems, and we suspect that this is the case here. The prediction losses are higher here than in our main results as reported in Table 2, at least in part because the between-session variance in cooperation is larger in the later supergames: the within-treatment variance in average cooperation is 0.0145 in the first halves of the sessions and 0.0196 in the second halves.[19]

6.2 Extrapolating to hypothetical session lengths

We generate predictions for the treatments in Dal Bó and Fréchette (2011), since these capture a nice range of behavior. For each of the treatments, 1,000 populations with 14 participants (which was the average in the original paper) were simulated for 10,000 supergames, with randomly drawn supergame lengths. We then simulated the learning model with the average (across folds) parameters estimated on the time

[19] To calculate these variances we restrict to treatments with at least 2 sessions.

path in table 4. Using these simulations we can compute the median level of average cooperation and its 90% confidence interval.

Δ^{RD}	δ	Q05	Median	Q95
-0.32	0.50	0.00	0.00	0.04
-0.11	0.50	0.00	0.00	0.05
0.11	0.50	0.00	0.43	0.79
-0.07	0.75	0.00	0.00	0.38
0.14	0.75	0.18	0.51	0.83
0.36	0.75	0.54	0.81	1.00

Table 8: Simulated cooperation after 10,000 supergames, 14 participants per session.

Δ^{RD}	δ	Q05	Median	Q95
-0.32	0.50	0.00	0.00	0.02
-0.11	0.50	0.00	0.00	0.02
0.11	0.50	0.15	0.44	0.64
-0.07	0.75	0.00	0.01	0.28
0.14	0.75	0.35	0.52	0.68
0.36	0.75	0.68	0.80	0.88

Table 9: Simulated cooperation after 10,000 supergames, 100 participants per session.

Tables 8 and 9 show quite wide 90% intervals for intermediate values of Δ^{RD}. This seems due to the effect of the randomness on experience and learning, which is more pronounced when the population is small. In the treatment $\Delta^{RD} = 0.11$, even after 10,000 supergames, the 90% interval goes from 0% to 79%. (With populations of 100 participants, the 90% interval is smaller but still substantial; it goes from 15% to 64%.) This randomness comes in part from random initial play in a finite population and in part from the randomness in the realized supergame lengths. Even if we increase the population size to 1,000, the 90% interval is still from 24% to 60%. However, if we also let all the simulated supergames have the expected number of rounds, the 90% interval is only 44% to 49%.

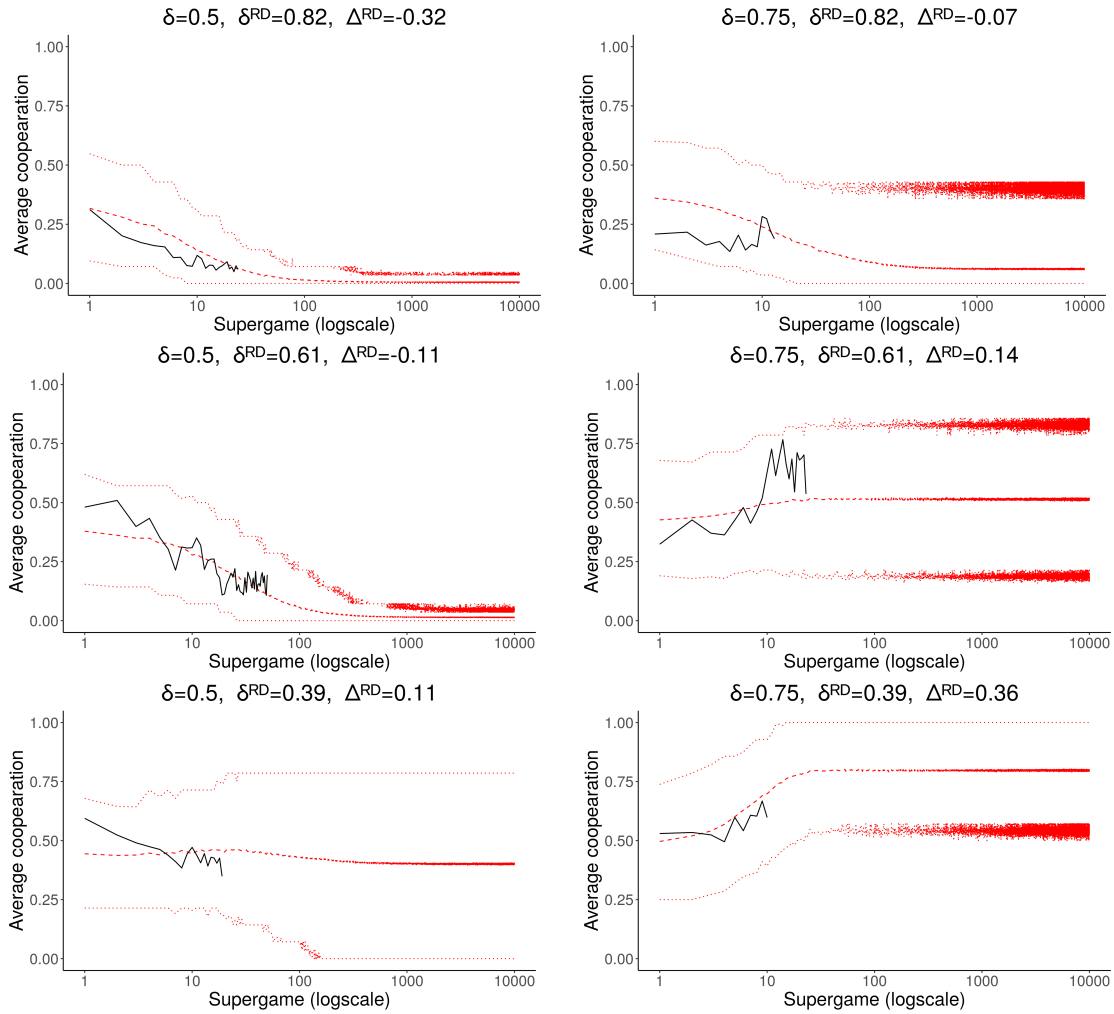

Figure 5: Predictions and actual behavior for six different treatments. The solid black line corresponds to the data, the red lines depict average cooperation and the middle 90% interval in 1,000 simulated populations.

Figure 5 displays the actual data and our confidence intervals for the time paths of cooperation in these treatments.[20]

The intervals are smaller in treatments where Δ^{RD} has a more extreme value in either direction. For $\Delta^{RD} < 0$, we predict less than 50% cooperation, and for $\Delta^{RD} = -0.32$ cooperation is almost certain to decrease. For $\Delta^{RD} = 0.14$ we see a slow increase in initial round cooperation to 51%, and for $\Delta^{RD} = 0.36$ we predict relatively

[20]Dal Bó and Fréchette (2011) use their pure-strategy belief learning model to produce similar plots for initial-round cooperation. Visual inspection suggests that our simpler model fits the data about as well.

fast and certain convergence to a high cooperation rate.

We get a broader picture of the long-run predictions by replicating this exercise for all 28 treatments in the data. In figure 6 we see the average cooperation after 10,000 supergames, predicted by simulating 1,000 populations of size 16 for each treatment. We see that for Δ^{RD} between 0 and 0.3, even after 10,000 supergames, the learning model does not predict either very high or very low rates of cooperation.

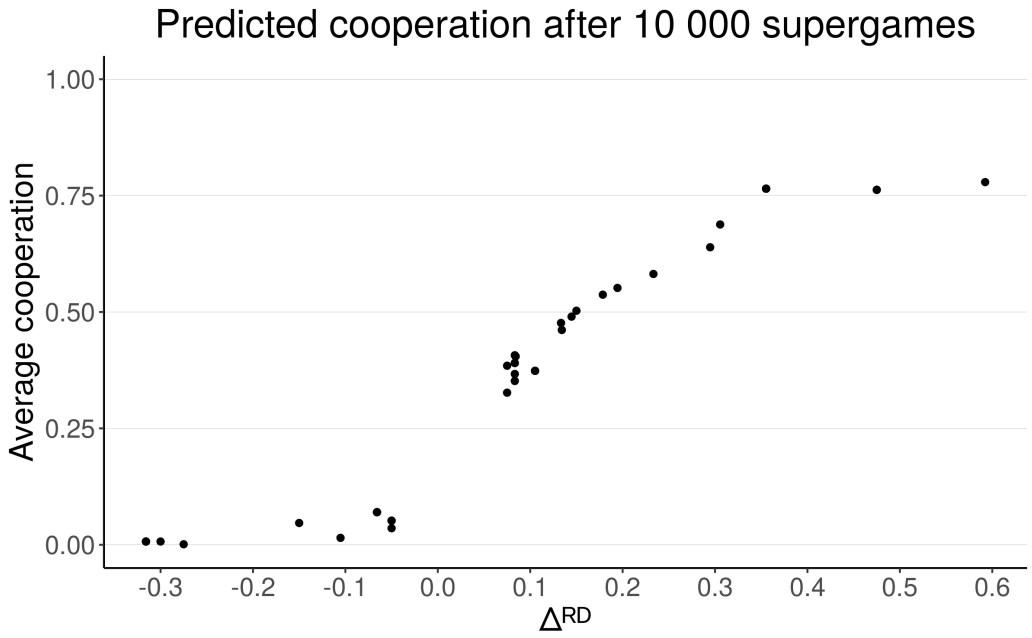

Figure 6: Predicted average cooperation after 10,000 supergames

7 Conclusion

The IRL-SG is simple and portable, with only 6 parameters and only one type of agent. It predicts average cooperation in a session by using a simulation to model the effect of learning on play in the initial round of each supergame, holding the strategy at non-initial round fixed. The model lets us capture the effect of playing more supergames on average cooperation, so that we can predict what average cooperation rates would be with longer lab sessions.

Our results show that the main way Δ^{RD} influences cooperation rates is through its effect on the probability of cooperation in the initial round of a match. Initial cooperation is positively reinforced when $\Delta^{RD} > .15$, so in these games the probability

of cooperating in the initial round increases over the course of a session. Initial cooperation is negatively reinforced when $\Delta^{RD} < 0$, so here initial cooperation rates drift down. For intermediate values of Δ^{RD}, a participant's overall payoff is about the same regardless of how they play in the initial round, which is why in these games initial cooperation rates stay roughly constant throughout a session.

In addition to explaining the influence of Δ^{RD}, the IRL-SG also captures the influence of additional determinants of cooperation, such as the composition of Δ^{RD}: It matters how a given Δ^{RD} is achieved. In particular, different values of $(g - l)$ have different implications for the same Δ^{RD}, and moreover the sign of the effect depends on the magnitude of Δ^{RD}. While there is not enough data to directly test these effects, it is a likely explanation to why the IRL-SG predicts cooperation better than just OLS on Δ^{RD}.

The analysis in our paper requires data from a great many sessions and treatments, and it was only made possible by the work of the researchers whose data we used. For this reason it only considers the prisoner's dilemma with perfect monitoring. In repeated games with implementation errors or imperfect monitoring, people seem to use more complex strategies with longer memory (Fudenberg, Rand and Dreber, 2012). There are not yet enough experimental studies of these games to see if Δ^{RD} plays a significant role there, let alone to support the sort of analysis we do here. Once there are, it would be useful to extend our analysis of average cooperation rates to this case.

References

Aoyagi, M., V. Bhaskar, and G. Fréchette. 2019. "The Impact of Monitoring in Infinitely Repeated Games: Perfect, Public, and Private." *American Economic Journal: Microeconomics*, 11: 1–43.

Athey, S., and K. Bagwell. 2001. "Optimal Collusion with Private Information." *RAND Journal of Economics*, 32: 428–465.

Backhaus, T., and Y. Breitmoser. 2020. "God Does Not Play Dice, but Do We?" *CRC TRR 190 #96*. http://hdl.handle.net/10419/185766.

Blonski, M., and G. Spagnolo. 2015. "Prisoners' other Dilemma." *International Journal of Game Theory*, 44: 61–81.

Blonski, M., P. Ockenfels, and G. Spagnolo. 2011. "Equilibrium Selection in the Repeated Prisoner's Dilemma: Axiomatic Approach and Experimental Evidence." *American Economic Journal: Microeconomics*, 3: 164–92.

Breitmoser, Y. 2015. "Cooperation, but No Reciprocity: Individual Strategies in the Repeated Prisoner's Dilemma." *American Economic Review*, 105: 2882–2910.

Camerer, C., and T.-H. Ho. 1999. "Experience-weighted attraction learning in normal form games." *Econometrica*, 67: 827–874.

Chassang, S. 2010. "Fear of Miscoordination and the Robustness of Cooperation in Dynamic Global Games with Exit." *Econometrica*, 78: 973–1006.

Cheung, Y.-W., and D. Friedman. 1997. "Individual Learning in Normal Form Games: Some Laboratory Results." *Games and Economic Behavior*, 19: 46–76.

Dal Bó, P. 2005. "Cooperation Under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games." *American Economic Review*, 95: 1591–1604.

Dal Bó, P., and G. Fréchette. 2011. "The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence." *American Economic Review*, 101: 411–29.

Dal Bó, P., and G. Fréchette. 2018. "On the Determinants of Cooperation in Infinitely Repeated Games: A Survey." *Journal of Economic Literature*, 56: 60–114.

Dal Bó, P., and G. Fréchette. 2019. "Strategy Choice in the Infinitely Repeated Prisoner's Dilemma." *American Economic Review*, 109: 3929–52.

Engle-Warnick, J., and R. Slonim. 2006. "Learning to Trust in Indefinitely Repeated Games." *Games and Economic Behavior*, 54: 95–114.

Erev, I., and A. Roth. 1998. "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria." *American Economic Review*, 88: 848–81.

Erev, I., and A. Roth. 2001. "Simple Reinforcement Learning Models and Reciprocation in the Prisoner's Dilemma Game." *Bounded Rationality: The Adaptive Toolbox*, ed. Gerd Gigerenzer and Reinhard Selten, Chapter 12, 215–231. The MIT Press.

Fudenberg, D., and A. Liang. 2019. "Predicting and Understanding Initial Play." *American Economic Review*, 109: 4112–41.

Fudenberg, D., D. Rand, and A. Dreber. 2012. "Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World." *American Economic Review*, 102: 720–49.

Fudenberg, D., J. Kleinberg, A. Liang, and S. Mullainathan. 2021. "Measuring the Completeness of Economic Models." *Journal of Political Economy*, Forthcoming.

Hanaki, N., R. Sethi, I. Erev, and A. Peterhansl. 2005. "Learning Strategies." *Journal of Economic Behavior & Organization*, 56: 523–542.

Harrington, J. 2017. *The Theory of Collusion and Competition Policy.* Cambridge:The MIT Press.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning. Springer Series in Statistics*, New York, NY:Springer.

Honhon, D., and K. Hyndman. 2020. "Flexibility and Reputation in Repeated Prisoner's Dilemma Games." *Management Science*, 66: 4998–5014.

Ioannou, C., and J. Romero. 2014. "A Generalized Approach to Belief Learning in Repeated Games." *Games and Economic Behavior*, 87: 178–203.

Kruskal, W. 1987. "Relative Importance by Averaging over Orderings." *The American Statistician*, 41: 6–10.

Lipovetsky, S. 2006. "Entropy Criterion in Logistic Regression and Shapley Value of Predictors." *Journal of Modern Applied Statistical Methods*, 5: 95–106.

Lundberg, S., and S.-I. Lee. 2017. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems*, 30: 4765–4774.

Mengel, F., S. Weidenholzer, and L. Orlandi. 2021. "Match Length Realization and Cooperation in Indefinitely Repeated Games." https://dx.doi.org/10.2139/ssrn.3777155.

Mishra, S. 2016. "Shapley Value Regression and the Resolution of Multicollinearity." https://dx.doi.org/10.2139/ssrn.2797224.

Molnar, C. 2019. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.* https://christophm.github.io/interpretable-ml-book/.

Proto, E., A. Rustichini, and A. Sofianos. 2019. "Intelligence, Personality, and Gains from Cooperation in Repeated Interactions." *Journal of Political Economy*, 127: 1351–1390.

Rand, D., and M. Nowak. 2013. "Human Cooperation." *Trends in Cognitive Sciences*, 17: 413–425.

Romero, J., and Y. Rosokha. 2018. "Constructing Strategies in the Indefinitely Repeated Prisoner's Dilemma Game." *European Economic Review*, 104: 185–219.

Romero, J., and Y. Rosokha. 2019. "A Model of Adaptive Reinforcement Learning." *Available at SSRN 3350711.*

Rotemberg, J., and G. Saloner. 1986. "A Supergame-Theoretic Model of Price Wars during Booms." *American Economic Review*, 76: 390–407.

Wright, J., and K. Leyton-Brown. 2017. "Predicting Human Behavior in Unrepeated, Simultaneous-move Games." *Games and Economic Behavior*, 106: 16–37.

A Numerical Estimation of Learning Models

To simulate a decision, a number $r \sim Uniform(0, 1)$ is drawn, and if that number is lower than the probability of cooperation for the simulated individual, she cooperates, otherwise defects. Similarly, the type of each individual is decided by a random draw. By fixing the draws of these values r, we get a deterministic function.

The resulting function is locally flat, which means that finding an optimum is difficult. To address this problem we first generate 30 candidate points using the following global differential evolution[21] optimization in parallel, using 100 individuals with a common set of random numbers.

1. First a population is initialized: For each agent x, we pick 3 new agents a, b, c from the population of candidates and generate a new candidate x'. Each parameter x_i of x is updated with some probability CR (the cross-over probability), and if it is updated the new value is given by $x'_i = a_i + F * (b_i - c_i)$. Once this is done, we compare the new value $f(x')$ with the old $f(x)$. If the this results in a lower loss, the new candidate replaces the old in the population, and otherwise it is thrown away.

2. After a fixed amount of time, the best candidate from this algorithm is used as a starting point for a Nelder-Mead algorithm that performs a local, gradient-free, optimization, using a different fixed realization of the random variables. The output of this local optimization is then returned as a candidate solution.

Once these 30 candidate points are found, they are each evaluated using a population size of 3,000, with a new fixed realization of the random draws r for all 30 candidates. The best of these parameters are then returned as the solution.

[21] From the package BlackBoxOptim.jl.

B Alternative models

Here we describe variations of the IRL-SG model that did not improve predictions.

IRL-SG with a recency effect. To allow more recent supergames to have a larger impact on behavior, we consider a model with experience weighted by recency. In that model, experience is updated according to

$$e_i(s) = \lambda \cdot a_i(s-1) \cdot V_i(s-1) + \rho \cdot e_i(s-1),$$

where $\rho \in [0, 1]$ discounts previous experiences.

Flexible reinforcement threshold. In the baseline IRL-SG, an initial action that leads to a positive payoff is reinforced. Here we add a parameter τ to relax this, so that experience is updated according to

$$e_i(s) = \lambda \cdot a_i(s-1) \cdot (V_i(s-1) - \tau) + \cdot e_i(s-1).$$

In other words, initial round actions are reinforced with respect to the difference with the threshold τ and not 0.[22]

The IRL-SG and AllD. This model adds a type which plays AllD with a fixed error ε, increasing the number of parameters by 2.

Learning with memory-1. Here we drop the semi-grim requirement that $\sigma_{DC} = \sigma_{CD}$, increasing the total number of parameters to 7.

Learning with flexible memory-1. In the next model, we allow these memory-1 behaviors to depend on Δ^{RD}. In this case we have

$$\sigma_h = \frac{1}{1 + \exp(-(\alpha^h + \beta^h \cdot \Delta^{RD}))}$$

In total this model has 11 parameters, but allows for the possibility that people, for example, cooperate more after a DC history if Δ^{RD} is high.

[22] We also considered several other variations where the threshold depended on for example the expected length of the supergame or Δ^{RD}; none of these did did better than this simple threshold model.

B.1 Results

Table 10: Out-of-sample prediction loss (MSE) for per-session average cooperation

Model	Avg C MSE	S.E.	Improvement
Constant	0.0517	(0.0040)	-
OLS on (δ, g, l)	0.0196	(0.0024)	62.1%
OLS on Δ^{RD}	0.0189	(0.0020)	63.4%
OLS	0.0153	(0.0017)	70.4%
GBT	0.0152	(0.0016)	70.6%
Lasso	0.0146	(0.0016)	71.7%
SVR	0.0145	(0.0016)	71.9%
Pure strategy belief learning w/o trembles	0.0229	(0.0025)	55.7%
Pure strategy belief learning w/ trembles	0.0191	(0.0020)	63.0%
Pure strategy reinf. learning w/ trembles	0.0175	(0.0020)	66.1%
IRL-SG and AllD	0.0143	(0.0015)	72.3%
IRL-SG with recency	0.0143	(0.0015)	72.3%
IRL-SG with flexible threshold	0.0141	(0.0015)	72.7%
Learning at all memory-1, two rates	0.0141	(0.0016)	72.7%
Learning with flexible memory-1	0.0140	(0.0016)	72.9%
Learning with memory-1	0.0139	(0.0015)	73.1%
Learning at all memory-1	0.0139	(0.0016)	73.1%
IRL-SG	0.0138	(0.0015)	73.3%
IRL-SG, two types	0.0137	(0.0015)	73.5%

Table 11: Out-of-sample prediction loss (MSE) for the time path of cooperation

Model	Time-path MSE	S.E.	Improvement
Constant	0.0775	(0.0050)	-
OLS on (δ, g, l)	0.0403	(0.0023)	48.0%
OLS on Δ^{RD}	0.0398	(0.0025)	48.6%
OLS	0.0324	(0.0019)	58.2%
SVR	0.0324	(0.0019)	58.2%
Lasso	0.0323	(0.0019)	58.3%
GBT	0.0321	(0.0020)	58.6%
Pure strategy belief learning w/o trembles	0.0435	(0.0030)	43.9%
Pure strategy reinf. learning w/ trembles	0.0395	(0.0026)	49.0%
Pure strategy belief learning w/ trembles	0.0377	(0.0025)	51.3%
Learning at all memory-1 two rates	0.0322	(0.0021)	58.4%
IRL-SG with recency	0.0319	(0.0020)	58.8%
Learning at all memory-1	0.0316	(0.0022)	59.2%
IRL-SG with threshold	0.0313	(0.0020)	59.6%
Learning with flexible memory-1	0.0312	(0.0020)	59.7%
IRL-SG and AllD	0.0311	(0.0019)	59.9%
Learning with memory-1	0.0310	(0.0020)	60.0%
IRL-SG	0.0309	(0.0020)	60.1%
IRL-SG, two types	0.0303	(0.0019)	60.9%

C Game parameters and learning

Our learning model assumes that game parameters and experience influence initial round cooperation through the sum $\alpha + \beta \cdot \Delta^{RD} + e_i(s)$. We can thus interpret $\alpha + \beta \cdot \Delta^{RD}$ as the direct effect of the game parameters and $e_i(s)$ as the direct effect of learning. We here try to answer how much of the behavior is directly driven by learning and how much is driven by the game parameters, according to our learning model. Since these two values enter the expression in the same way, they are directly comparable.

We consider the last supergame of each experimental session. We consider the actual data and a simulated data set with 16 participants in each session. When we consider the actual data for an individual, we look at the initial round actions they took and their observed realized, and calculate the corresponding value for $e_i(s)$ in the last supergame. For the simulated data, we instead simulate the whole sequence

of play, and use the simulated values to calculate $e_i(s)$.

To get a numerical estimate of the relative importance we can look at how much of the variation in predicted initial round cooperation is driven by the two effects. The total average variance in initial round cooperation is given by

$$\text{Var}(p|e, \Delta^{RD}) = \sum_{i \in I} \left(\frac{1}{1 - \exp\left(-(\alpha + \beta \Delta^{RD} + e_i(s)\right)} - \overline{p} \right)^2 / |I|$$

where I is the set of all individuals, and $\overline{p}$ is the average predicted initial round cooperation. We can compare this to the variation in predicted cooperation from the direct learning effect and the direct game parameter effect respectively.

$$\text{Var}(p|\Delta^{RD}) = \sum_{i \in I} \left(\frac{1}{1 - \exp\left(-(\alpha + \beta \Delta^{RD})\right)} - \overline{p}(\Delta^{RD}) \right)^2 / |I|$$

$$\text{Var}(p|e) = \sum_{i \in I} \left(\frac{1}{1 - \exp\left(-e_i(s)\right)} - \overline{p}(e) \right)^2 / |I|.$$

To calculate the relative importance of Δ^{RD} we compare the incremental variance introduced by Δ^{RD} to the fraction of the variance introduced by $e_i(s)$, averaged over which term we add first, and dived by the total variance

$$\text{Relative Importance}(\Delta^{RD}) = \frac{\text{Var}(p|\Delta^{RD}) + \Big(\text{Var}(p|e, \Delta^{RD}) - \text{Var}(p|e)\Big)}{2} / \text{Var}(p|e, \Delta^{RD})$$

$$\text{Relative Importance}(e) = \frac{\text{Var}(p|e) + \Big(\text{Var}(p|e, \Delta^{RD}) - \text{Var}(p|\Delta^{RD})\Big)}{2} / \text{Var}(p|e, \Delta^{RD}).$$

This is the relative Shapley value of the two effects.[23] It can be calculated on either the individual or treatment level, where the probabilities $p_i(s)$ are first averaged for each session.

[23](Kruskal, 1987) and Mishra (2016) use the Shapley value to analyze regressions, (Lipovetsky, 2006) uses them for logistic regressions, and (Lundberg and Lee, 2017; Molnar, 2019) for general machine learning algorithms.

Data	$\text{Var}(p\|e, \Delta^{RD})$	$\text{Var}(p\|e)$	$\text{Var}(p\|\Delta^{RD})$	Rel Imp $e_i(s)$	Rel Imp Δ^{RD}
Simulated individual	0.195	0.188	0.006	96.8%	3.2%
Actual individual	0.185	0.18	0.005	97.2%	2.8%
Simulated treatment	0.059	0.052	0.005	89.5%	10.5%
Actual treatment	0.055	0.046	0.004	87.7%	12.3%

Table 12: Relative importance measures.

The table shows that in our model experience drives most of the variation initial round cooperation: In both the simulated and actual data, $e_i(s)$ is responsible for roughly 97% of the variation in predicted individual behavior, and roughly 88% of the variation in predicted initial round cooperation between treatments.

D Pairwise tests

Here we consider paired differences between the IRL-SG and the different alternatives considered. With the 10 different 10-fold cross-validation splits, each session is predicted 10 times. This way we capture the randomness that comes from having a particular assignment to different folds.

Since the same 10 folds are used for all the models, we can bootstrap from these 1610 differences to produce more reliable estimates of the differences between models. The tables below show bootstrapped standard errors and bootstrapped 95% confidence intervals for the paired differences between IRL-SG and alternatives.

Table 13 shows that learning with semi-grim is significantly better than OLS on Δ^{RD} and all pure strategy learning models. Furthermore, we see that the differences for the generalizations are in general not significant and neither are the differences with the ML predictions.

Table 14 shows that the picture is similar for the time-path prediction problem.

Model	Difference	S.E.	C.I.
OLS on Δ^{RD}	-0.0050	(0.0014)	[-0.0079, -0.0022]
GBT	-0.0013	(0.0011)	[-0.0035, 0.0008]
Lasso	-0.0007	(0.0009)	[-0.0024, 0.0010]
SVR	-0.0006	(0.0010)	[-0.0025, 0.0013]
Pure strategy belief learning w/o trembles	-0.0090	(0.0021)	[-0.0133, -0.0051]
Pure strategy belief learning w/ trembles	-0.0053	(0.0016)	[-0.0085, -0.0021]
Pure strategy reinf. learning w/ trembles	-0.0037	(0.0015)	[-0.0066, -0.0009]
IRL-SG and AllD	-0.0004	(0.0005)	[-0.0015, 0.0006]
IRL-SG with recency	-0.0004	(0.0005)	[-0.0015, 0.0006]
IRL-SG with threshold	-0.0002	(0.0004)	[-0.0009, 0.0004]
Learning with flexible memory-1	-0.0002	(0.0006)	[-0.0013, 0.0009]
Learning at all memory-1, two rates	-0.0002	(0.0006)	[-0.0015, 0.0010]
Learning with memory-1	-0.0001	(0.0004)	[-0.0008, 0.0007]
Learning at all memory-1	-0.0000	(0.0006)	[-0.0012, 0.0011]
IRL-SG, two types	0.0002	(0.0005)	[-0.0008, 0.0012]

Table 13: Paired differences and 95% confidence intervals with the IRL-SG for the average cooperation prediction task; negative values indicate that the IRL-SG has lower MSE.

Model	Difference	S.E.	C.I.
OLS on Δ^{RD}	-0.0089	(0.002)	[-0.0129, -0.0049]
SVR	-0.0015	(0.0012)	[-0.0039, 0.0009]
Lasso	-0.0014	(0.0012)	[-0.0037, 0.0010]
GBT	-0.0011	(0.0014)	[-0.0038, 0.0015]
Pure strategy belief learning without trembles	-0.0126	(0.0023)	[-0.0173, -0.0082]
Pure strategy reinf. learning with trembles	-0.0086	(0.0018)	[-0.0122, -0.0050]
Pure strategy belief learning with trembles	-0.0068	(0.0021)	[-0.0110, -0.0028]
Learning at all memory-1 two rates	-0.0013	(0.0008)	[-0.0028, 0.0003]
IRL-SG with recency	-0.0010	(0.0008)	[-0.0025, 0.0005]
Learning at all memory-1	-0.0007	(0.0008)	[-0.0023, 0.0009]
IRL-SG with threshold	-0.0004	(0.0005)	[-0.0013, 0.0007]
Learning with flexible memory-1	-0.0003	(0.0007)	[-0.0018, 0.0011]
Learning with memory-1	-0.0001	(0.0004)	[-0.0009, 0.0008]
IRL-SG and AllD	-0.0001	(0.0006)	[-0.0012, 0.0010]
IRL-SG, two types	0.0007	(0.0007)	[-0.0007, 0.0020]

Table 14: Paired differences and 95% confidence intervals with the IRL-SG for the time-path prediction task; negative values indicate that the IRL-SG has lower MSE.

Table 15 shows that the IRL-SG is significantly better than the atheoretical prediction algorithms at extrapolating from short to long sessions.

Model	Difference	S.E.	C.I.
Constant	-0.0475	(0.0054)	[-0.0582, -0.0371]
Lasso	-0.0065	(0.0028)	[-0.0121, -0.0010]
SVR	-0.0065	(0.0028)	[-0.0121, -0.0010]
OLS on Δ^{RD}	-0.0061	(0.0022)	[-0.0105, -0.0018]
GBT	-0.0046	(0.0020)	[-0.0086, -0.0006]

Table 15: Paired differences and 95% confidence intervals with the IRL-SG when estimating on 1st half and evaluating on 2nd half; negative values indicate that the IRL-SG has lower MSE.