Predicting Average Cooperation in the Repeated Prisoner's Dilemma*

Drew Fudenberg[1] and Gustav Karreskog[2]

[1]Department of Economics, MIT

[2]Department of Economics, Stockholm School of Economics

March 25, 2021

Abstract

We predict cooperation rates across treatments in the experimental play of the indefinitely repeated prisoner's dilemma using simulations of a simple learning model. We suppose that learning and the game parameters only influence play in the initial round of each supergame. 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 machine learning algorithms. Our results let us predict how cooperation rates change with longer experimental sessions, and explain and sharpen past findings on the role of strategic uncertainty.

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

*We thank Anna Dreber Almenberg, Mathias Blonski, Yves Breitmoser, Ying Gao, Annie Liang, Indira Puri, Emanuel Vespa, Jörgen Weibull, David Rand, and seminar participants at Goethe University Frankfurt, U Queensland, and UVA for helpful comments. NSF grants SES 1643517 and 1951056 provided financial support.

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 cooperation rates in the experimental play of the prisoner's dilemma and its exogenous parameters as a prediction problem. We formulate and evaluate 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.

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 these strategies, and moreover, that exactly half of the population uses each one. (This is not meant as a realistic assumption, but is simply an explanation of how the 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.

To make predictions with the model, we use simulations of play, and 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 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 generates more accurate predictions of both average cooperation and the time path of cooperation in a session better than any of the black-box methods we consider, or 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. Here we find that even in the very long run, high rates of cooperation are predicted only when its benefit is high compared to its risk. For intermediate values, substantial shares of initial round cooperators and initial round defectors coexist in the population.

To further evaluate our learning model, we then consider the problem of predicting the next action that a participant will play, which is closely related to the commonly-studied task of identifying the strategies used by the participants. We find that the naive rule that a player's next action will be the same as their previous one predicts quite well, and that our learning model predicts almost as well as more complicated models. However, in contrast to our primary prediction problems, allowing for a mixture of a few different types does improve predictions of the next action taken.

As we detail in Section 3, past work has already found evidence that most players use memory-1 strategies and that overall cooperation rates depend on Δ^{RD}. In our preliminary data analysis, we sharpen the latter conclusion by finding that 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. Moreover, according to our

model, the direct effect of game parameters on cooperation rates is much smaller than their indirect effect through learning. As a consequence, participants in a session can behave differently even if they follow the same learning model.

This simple learning model does not use individual characteristics as data, so there are regularities it cannot capture. Indeed, Proto, Rustichini and Sofianos (2019) show that intelligence, and to a lesser extent, other personality traits, affect how people play infinitely repeated games. However, the learning model is parsimonious and portable, and predicts average cooperation and its time path well.

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 ends 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).

	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)

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

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

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

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.

Applied theoretical work on repeated games often assumes 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, Grim is risk dominant in a 2x2 matrix game with the strategies Grim and Always Defect iff

$$\delta \geq (g+l)/(1+g+l) \equiv \delta^{\text{RD}}.$$

The composite parameter referred to above is the difference between the actual discount factor and this threshold:

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

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 $e_i(s)$ 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).

Thus initial behavior 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 $e_i(s)$.

To model learning, we suppose that after each supergame s, $e_i(s)$ 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 $a_i(s) \in \{-1, 1\}$ is the action taken, $V_i(s)$ is the total payoff received in supergame s, λ determines the strength of the learning, and $e_i(1) = 0$.[3] 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 5.4 we relax this assumption and consider multiple extensions, but this does not improve predictions.) In total, our model has 6 parameters, $(\alpha, \beta, \lambda, \sigma_{CC}, \sigma_{CD/DC}, \sigma_{DD})$.

Importantly, in our main analysis we do not make predictions based on the actual payoffs that participants received, but rather on simulations that suppose all participants used learning rules of the form (1) and (2). (We do use the actual payoffs when we turn to the problem of predicting the next action a given participant will play.)

3 Prior Work

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) show that the sign of Δ^{RD} is much more correlated

[3] We also tried an alternative specification where learning responds to the average payoff in a supergame instead of the total, but it performed less well. This suggests that learning between supergames is stronger when the supergames are longer.

with high cooperation rates than the sign of $(\delta - \delta^{\text{SPE}})$.[4]

Several papers estimate the strategies used by participants on the assumption that each participant uses a fixed strategy either in the entire session or in the latter part of it. A consistent finding in the papers that assume the use of pure strategies is 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). See for example Dal Bó and Fréchette (2011); Fudenberg, Rand and Dreber (2012); Dal Bó and Fréchette (2018). In Romero and Rosokha (2018*a*) and Dal Bó and Fréchette (2019), the pure strategies used are elicited from the participants instead of being estimated. Those studies confirm the finding that a small set of memory-1 strategies are enough to capture most of the strategies used.[5]

More recent studies find evidence for the use of constant memory-1 mixed strategies. Breitmoser (2015) finds that strategies of the form "semi-grim" better fit play after the initial round than pure strategies do. These strategies are defined by $\sigma_{CC} > \sigma_{CD} = \sigma_{DC} > \sigma_{DD}$. Backhaus and Breitmoser (2018) follows up on this analysis by more carefully considering alternative models and behavior in the initial round. The authors argue that a combination of AllD and semi-grim, with the mixture estimated treatment by treatment, best fits behavior, and that only initial round behavior responds to incentives. [6]

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) argue that δ has no apparent effect on behavior in the first supergame, but a substantial

[4]Dal Bó and Fréchette (2011) use the alternative measure $\frac{(1-\delta)l}{1-(1-\delta)(1+g-l)}$ as a regressor; it is very correlated with Δ^{RD}.

[5]Fudenberg, Rand and Dreber (2012) show that longer memories are used when the intended actions are implemented with noise and only the realized actions are observed.

[6]Romero and Rosokha (2018*b*) elicits memory-1 mixed strategies and finds that a mixture of elicited mixed and pure strategies matches behavior better than pure strategies. In their data, $\sigma_{CD} = .45, \sigma_{DC} = .35$, and they reject semi-grim's restriction that $\sigma_{CD} = \sigma_{DC}$; in our larger data set $\sigma_{CD} = .31, \sigma_{DC} = .33$.

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. Dal Bó and Fréchette (2011) considers a simple belief learning model involving only TFT and AllD. [7]

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 larger literature on learning in games 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), who included experiments on the repeated prisoner's dilemma with perfect monitoring, deterministic payoffs, and constant parameters within a session from papers that were

[7] Erev and Roth (2001), Hanaki et al. (2005) and Ioannou and Romero (2014) study learning *within* supergames.

published before 2014. we consider only their treatments with $\delta > 0$. We 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)) increasing the number of observations by approximately 60%. Our resulting data set contains observations from 17 papers, 28 different treatments,[8] 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. In 20 of the sessions, $\delta < \delta^{\text{SPE}}$, so no cooperation can occur in a subgame perfect equilibrium. In 28, 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, $\delta > \delta^{\text{RD}}$.

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 and 53.6% for $\delta > \delta^{\text{RD}}$.

The average play after the different memory-1 histories, and the histories' frequencies, are shown in table 1. We see that the CD and DC histories are only a small subset of observations, roughly 15% 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 CD.

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

History	Avg C	N
CC	96.6%	59,435
CD	30.6%	16,705
DC	33.2%	16,707
DD	5.2%	74,621
initial	47.1%	64,830
Total	44.1%	232,298

[8]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.

To visualize how behavior differs depending on Δ^{RD}, we put the sessions into five groups: $\delta < \delta^{SPE}$, $\delta^{SPE} < \delta < \delta^{RD}$, $0 < \Delta^{RD} < 0.15$, $0.15 < \Delta^{RD} < 0.3$, and $0.3 < \Delta^{RD}$.

Here the first 2 groups were motivated by theory, while the subdivision of the treatments with $\Delta^{RD} > 0$ was based on the data. The thresholds and relative frequencies of Δ^{RD} can be seen in figure 2.

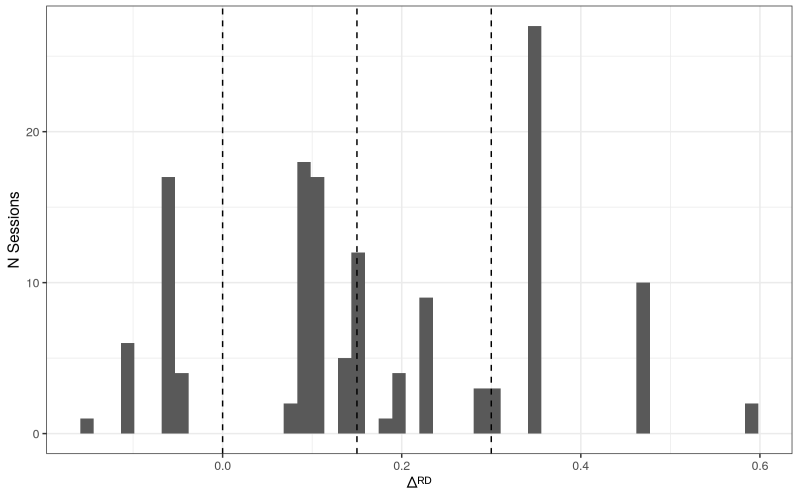

Figure 2: Distribution of Δ^{RD} for $\delta > \delta^{SPE}$

Figure 3 shows the evolution of cooperation during the first 10 supergames, restricted to sessions of at least 10 supergames (134 of 161), and in figure 4 the first 20 supergames restricted to the sessions that included at least 20 supergames (93 of 161).

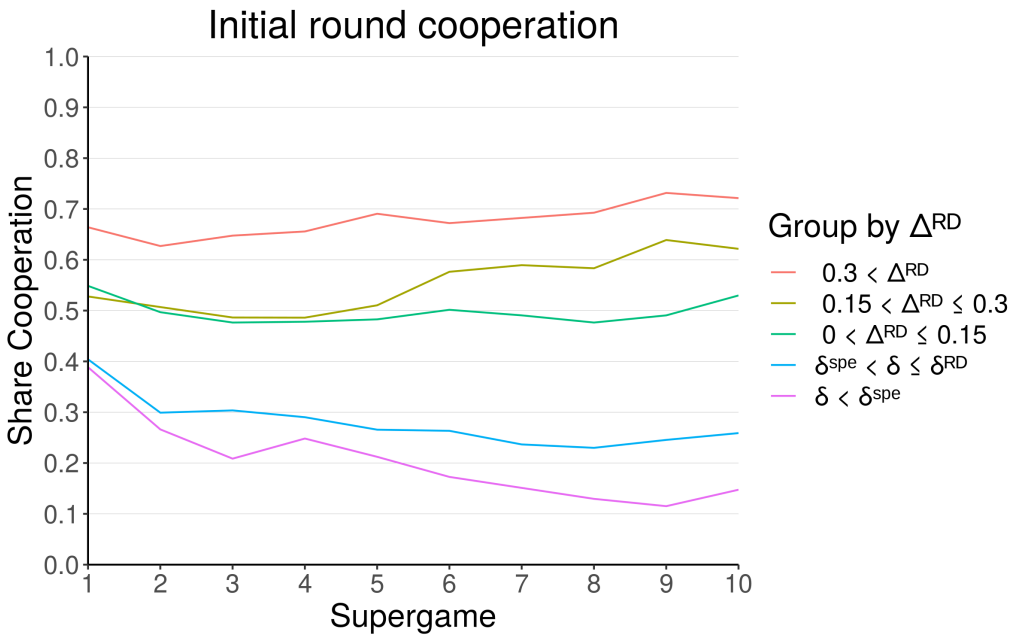

Figure 3: Cooperation in the initial round over the 10 first supergames.

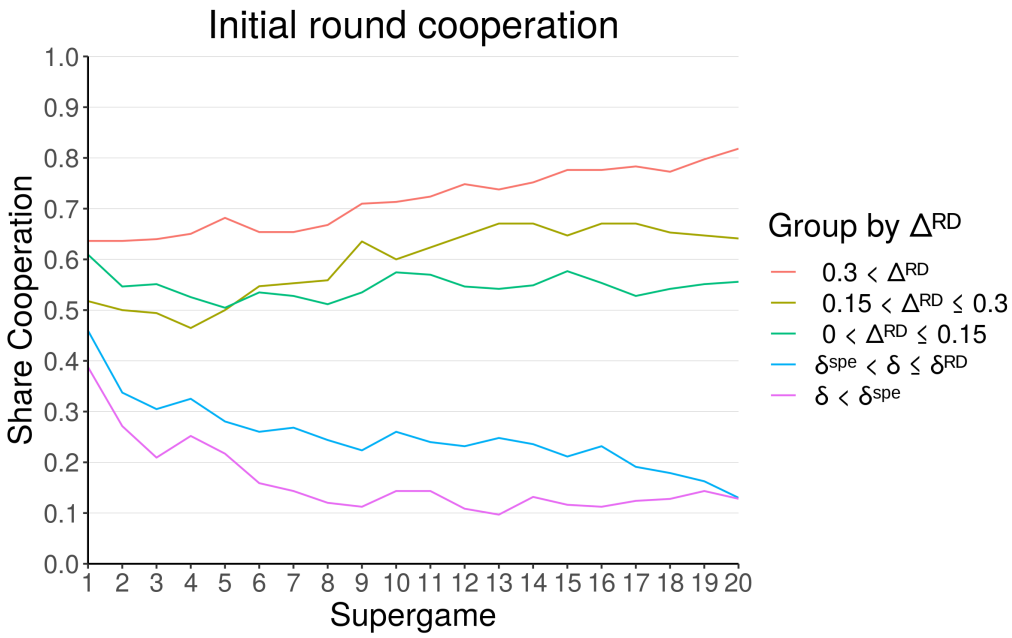

Figure 4: Cooperation in the initial round over the 20 first supergames.

The average rate of cooperation in the initial round of the supergames differs across the treatment groups. For $\Delta^{RD} > 0.15$ cooperation rates increase over the course of a session while for $\Delta^{RD} < 0$ they decrease. For sessions where cooperation is only marginally risk dominant, $0 < \Delta^{RD} < 0.15$, cooperation rates remain roughly constant at around 50%.

As we see in Online Appendix A, 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. As a further illustration of this, we look at the average cooperation in the non-initial rounds of the supergames. 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.

Summing up, the outcome of the initial round is highly predictive of the cooperation in the rest of the supergame, and taking into account the game parameters does not substantially improve these predictions. Similarly, we will find that the fit of our simple learning model is not improved by allowing the cooperation probabilities at non-initial histories to depend on game parameters. We find this somewhat surprising, but do not have a good explanation.[9] It is possible that the reason the game parameters seem to only matter in the initial rounds is somehow driven by selection or interaction effects that we fail to correct for, or that adding an additional, non-linear function of game parameters to Δ^{RD} would improve predictions, but none of these possibilities seem likely to eliminate the striking impact of initial play.

[9]Backhaus and Breitmoser (2018) also find that play after the initial round is relatively insensitive to the game parameters.

5 Predicting Cooperation

Our goal in this paper is to develop models that can successfully predict cooperation levels in repeated game experiments. 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 for us 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. [10]

5.1 Predicting Cooperation with Learning

To make predictions with the learning model, we simulate populations playing the different sessions assuming they behave according to the learning model. We make predictions using only the game parameters and the sequence of supergame lengths. In particular, we use the simulations to generate the experience levels e_i, and do not use data on the payoffs that people actually received in the sessions. 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 the first supergame, the individual experiences are updated according to equation (2), using the simulated values. The learning thus takes place between supergames. 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

[10] This use of a simple prediction rule as a benchmark is inspired by the completeness measure of (Fudenberg et al., 2020) but we do not have enough data to estimate the problem's irreducible error as the completeness measure requires.

exactly the same sequence of supergames as in the experimental session, updating the experience $e_i(s)$ 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 as predictions and compute the approximate prediction losses and associated standard errors.

We estimate the learning model 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 D we evaluate this estimation procedure on data simulated under different assumptions about how people actually behave, and confirm that it should work.

Our main learning model, presented earlier in equations (1) and (2), assumes all agents use the same learning rule, which is an oversimplification. In particular, past work has shown that in most experiments there is a non-negligible share of people who defect all or almost all of the time. As we show in section 7, adding a share of such individuals improves the prediction of the next action played, but as it does not improve predictions of the overall average cooperation rates, we do not include them in the estimates we report here.

The restriction to memory-1 strategies is motivated by past work and also by our machine learning analysis in Section 7. The assumption that play across treatments is the same except in the initial rounds is motivated by the descriptive statistics. Specifically, we assume that play at each non-initial memory-1 history follows the same semi-grim strategy.

We relax the assumption of fixed behavior at non-initial histories in section 5.4, which considers a more richly-parameterized model that lets play at these histories depend on Δ^{RD}. We also consider a model that extends learning to those non-initial histories. Neither of these extensions improve predictions.

5.2 Results

We now compare the performance of our learning model to that of OLS, Lasso, and Gradient Boosting Trees (GBT). 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, the sequence of supergame lengths, an indicator variable for whether $\Delta^{RD} > 0$, the difference between expected and realized supergame lengths in the first third of the session, the total difference between expected and realized supergame lengths, and some interaction terms. [11]

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.

In table 2 the out-of-sample prediction errors for average cooperation are shown for our learning model, a linear function of Δ^{RD}, and the best-performing atheoretical prediction method (in this case, Lasso). We see that our learning model is in fact better than the atheoretical prediction algorithms given the features we let the algorithms use. We also see that a linear function of Δ^{RD} is a strong predictor of behavior. We will return to question of the relationship of Δ^{RD} and average cooperation in subsection 5.3.

[11] In contrast to when we make predictions with learning models, we here directly predict average cooperation. The alternative method of training the ML algorithms on time paths, and using time path predictions to generate predictions of average cooperation yields very similar but slightly worse results.

Model	Avg C MSE	S.E.	Relative Improvement
Constant	0.0517	(0.0011)	-0.01%
OLS on Δ^{RD}	0.0189	(0.0006)	63.44%
Lasso	0.0145	(0.0005)	71.95%
Learning with semi-grim	0.0139	(0.0005)	73.11%

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

To estimate the out-of-sample prediction errors, we use 10-fold cross validation. This means we divide the sessions into 10 different folds. We split the data 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.[12] To estimate the standard errors of the estimated mean squared error (MSE), we do 10 different such cross validations, leading to a total of 100 MSEs estimated on different folds. Using these 100 different values, we estimate the standard errors of the out-of-sample MSE prediction error. By using the same folds for all models we can perform pairwise comparisons. Pairwise tests are presented in appendix section D. According to those pairwise tests, our learning model is indeed significantly better than the atheoretical prediction algorithms.[13]

At first glance, it might seem surprising that our learning model outperforms the ML algorithms. Part of the explanation is that our data set is relatively small by machine learning standards. In addition, our learning model is better able to incorporate the effect of the realized supergame lengths, as can be seen by what happens when we redo the estimations without using the realized supergame lengths as data. This increases the out-of-sample MSE for both the learning model and the best performing ML algorithm to 0.0158.

Not only is our proposed learning model better than alternatives at predicting

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

[13]The same data is used to estimate the model multiple times, so there are no asymptotic guarantees that these standard errors will match the true standard errors. We also consider non-parametric pairwise tests.

average cooperation in a session, it is also better at predicting the time path of cooperation: We better predict not only how much the participants cooperate on average, but how behavior evolves across and within supergames.

In table 3 we see similar results for predicting the time-path of cooperation.

Model	Time-path MSE	S.E.	Relative Improvement
Constant	0.0770	(0.0014)	0.0%
OLS on Δ^{RD}	0.0399	(0.0007)	48.18%
GBT	0.0322	(0.0006)	58.18%
Learning with semi-grim	0.0310	(0.0006)	59.74%

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

Figure 5 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, 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 the intermediate values of Δ^{RD}.

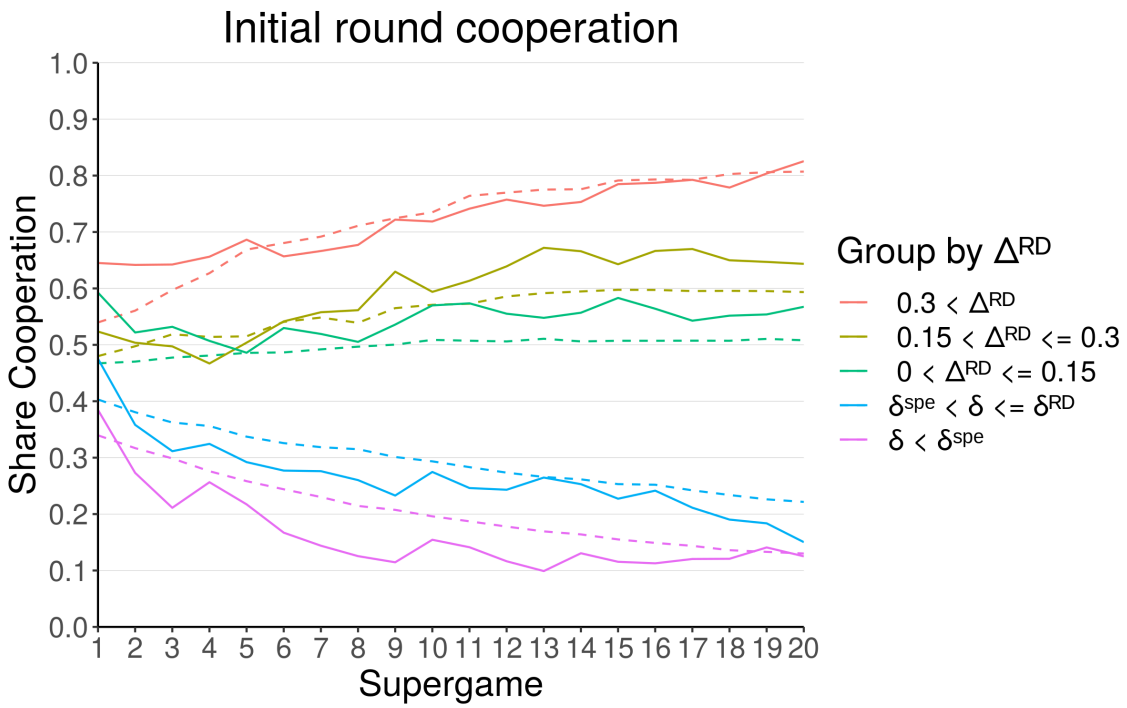

Figure 5: Actual (solid line) and out-of-sample predicted (dashed line) initial-round cooperation by supergame for sessions of at least 20 supergames .

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

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

which then enters into the probability of initial round cooperation by

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

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 supergame increases to 46.5%.

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 cooperation $p_i^{initial}(2)$ goes from 43.3% to 65.6%. An individual j experiencing DC in the first round and DD in the remaining 4 rounds gets payoff 3, which implies that $p_j^{initial}(2)$ would go down 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. 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 and not the direct influence of the game parameters.

5.3 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. The model's assumption that all individuals use the semi-grim strategy implies that higher rates of initial cooperation lead to more cooperation in that supergame, and the reinforcement-learning component of the model implies that this will lead to more cooperation in subsequent supergames.

To better understand the success of our learning model, and why Δ^{RD} is such a strong predictor, we relate the supergame payoffs participants receive to their initial actions. For each session ζ, let $\pi(C)$ be the average supergame payoff received by

participants who cooperated in the initial round and define $\pi(D)$ analogously. Figure 6 demonstrates the correlation between $\pi(C) - \pi(D)$ and Δ^{RD} in the data.

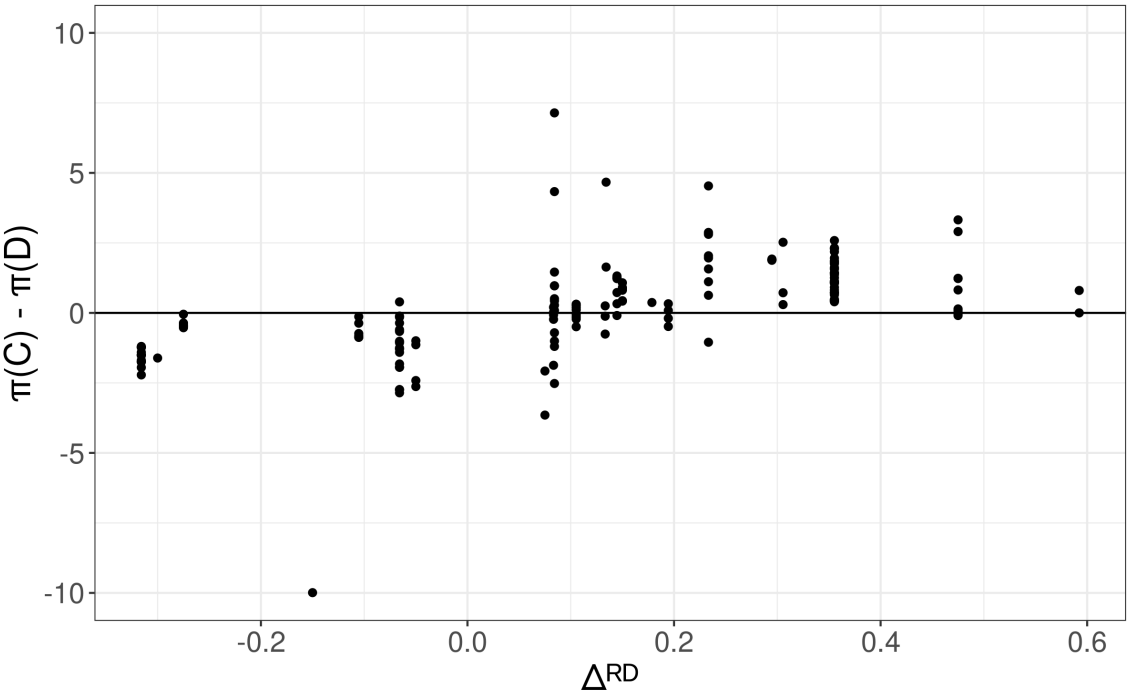

Figure 6: Average empirical difference between total payoff in supergames where the participant cooperated and defected. Each dot corresponds to one session.

For $\Delta^{RD} < 0$, defection is reinforced more strongly than cooperation in all but 1 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 $0 < \Delta^{RD} < 0.15$. In figure 7 we do the same analysis on simulated data. We simulate 100 participants for each session, playing the same sequence of supergame lengths as in the actual data, and calculate the corresponding value for $\pi(C) - \pi(D)$. The payoff difference has less variation due to the larger number of simulated participants than real participants, but it follows the same pattern as in the actual data.

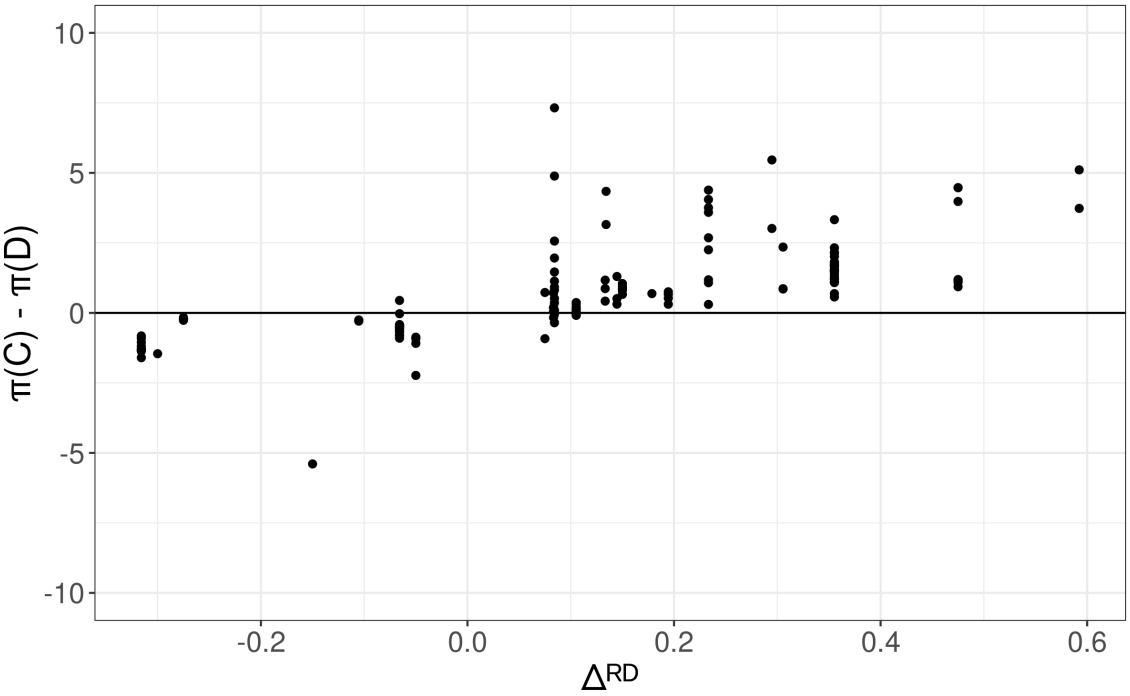

Figure 7: Average simulated difference between total payoff in supergames where the participant cooperated and defected. Each dot corresponds to one simulated session.

Even though $\pi(C) - \pi(D)$ is correlated with Δ^{RD}, we predict average cooperation better with our learning model than by 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.

5.4 Comparison with Alternative Models

We have now seen that our simple learning model can improve out-of-sample predictions of both average cooperation and the time path of cooperation. Here we want to answer two questions: Can we improve our model by adding more parameters or by introducing heterogeneity? Would a different learning model perform as well? In particular, we also want to see if a model that learns to play different pure strategies can at least match our model's performance.

Learning 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.

Parametric memory-1 behavior Our main learning model assumes that all players use a semi-grim mixed strategy. In in *Learning with memory-1* we drop the semi-grim requirement that $\sigma_{DC} = \sigma_{CD}$, increasing the total number of parameters to 7.

Flexible memory-1 behavior In the next step, we allow these memory-1 behaviors depend on Δ^{RD}. This model we call *Initial round learning with flexible memory-1*. 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.

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. We can extend the learning dynamic we have for the initial round to all memory-1 histories. The "full learning" model tracks 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. Experience at h is only updated when h occurs, and when it does, experience is updated for the rest of the supergame $V_i(t)$ according to

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

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

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

where $r(t)$ denotes the round at time t. This has the same number of parameters (11) as the previous model. 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 $\lambda_{memory-1}$.

Heterogeneous agents. To allow for heterogeneous agents, we now consider a mixture extension of our learning model. We assume that there are two different types with different parameters, and one variable deciding the share of the two types in the population. In sample, this of course improves predictions a little bit. Out of sample, however, there is at best a minor improvement.

When we consider the individual one-step ahead predictions, we will see that introducing heterogeneity does slightly improve predictions. One reason that we find so little evidence of type heterogeneity may be that the learning model with a single type has endogenous heterogeneity that can account for some of the observed heterogeneity: If an individual by chance defects in the initial round a few periods, they are likely to get a positive payoff in those supergames, thus reinforcing defection.[14] In contrast, adding a constant share of AllD players slightly decreases the accuracy

[14]Our data does not include individual characteristics such as gender, major, or cognitive ability.Proto, Rustichini and Sofianos (2019) find that more intelligent subjects are quicker to adjust their play to feedback.

of out-of-sample predictions.[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 AllD. Each participant has beliefs about how common TFT and AllD are in the population, which they update based (only) on opponents' moves in the initial rounds, and use 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 start 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

[15]This may be surprising in light of past findings that this form of heterogeneity is useful in predicting the next action played. See our discussion of this in Section 8.

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 learning with semi-grim. 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.

Model	Avg C MSE	S.E.	Relative Improvement
Pure strategy belief learning with trembles	0.0191	(0.0006)	63.05%
Pure strategy reinf. learning with trembles	0.0175	(0.0006)	66.15%
Learning with memory-1	0.0140	(0.0005)	72.92%
Learning at all h	0.0139	(0.0005)	73.11%
Learning with semi-grim	0.0139	(0.0005)	73.11%
Learning with semi-grim, two types	0.0136	(0.0005)	73.69%

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

Taken together, this suggests that our preferred, simple learning model 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.

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.

6.1 Extrapolating within observed sessions

Before we turn to 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 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 on the time paths in the first half of the session and use them to predict the average cooperation in the second half. To predict the average cooperation in the second half, we first predict the time path and calculate the resulting average cooperation. Table 6 displays the cross-validated MSE.

Model	Second half Session avg	Second half time path
Constant Prediction	0.066 (0.002)	0.081 (0.002)
Lasso	0.028 (0.001)	0.044 (0.001)
GBT:time-path	0.028 (0.001)	0.040 (0.001)
Learning with semi-grim	0.024 (0.001)	0.037 (0.001)

Table 6: 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. In principle this might be due to our particular ML implementations, but it is also true that atheoretical prediction algorithms can have trouble extrapolating to a 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.

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 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 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	Mean	Q95
-0.32	0.50	0.00	0.01	0.04
-0.11	0.50	0.00	0.01	0.05
0.11	0.50	0.00	0.40	0.79
-0.07	0.75	0.00	0.06	0.38
0.14	0.75	0.18	0.51	0.83
0.36	0.75	0.54	0.80	1.00

Table 7: Simulated cooperation after 10,000 supergames, 14 participants per session.

Δ^{RD}	δ	Q05	Mean	Q95
-0.32	0.50	0.00	0.01	0.02
-0.11	0.50	0.00	0.01	0.02
0.11	0.50	0.15	0.43	0.64
-0.07	0.75	0.00	0.05	0.28
0.14	0.75	0.35	0.52	0.68
0.36	0.75	0.68	0.79	0.88

Table 8: Simulated cooperation after 10,000 supergames, 100 participants per session.

Tables 7 and 8 show quite wide 90% intervals for intermediate values of Δ^{RD} in figure 8 due to the randomness of behavior and small population size. In the treatment $\Delta^{RD} = 0.11$, even after 10,000 supergames the 90% interval goes from 0% to 79%, and the average is just 40%. (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 also 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%.

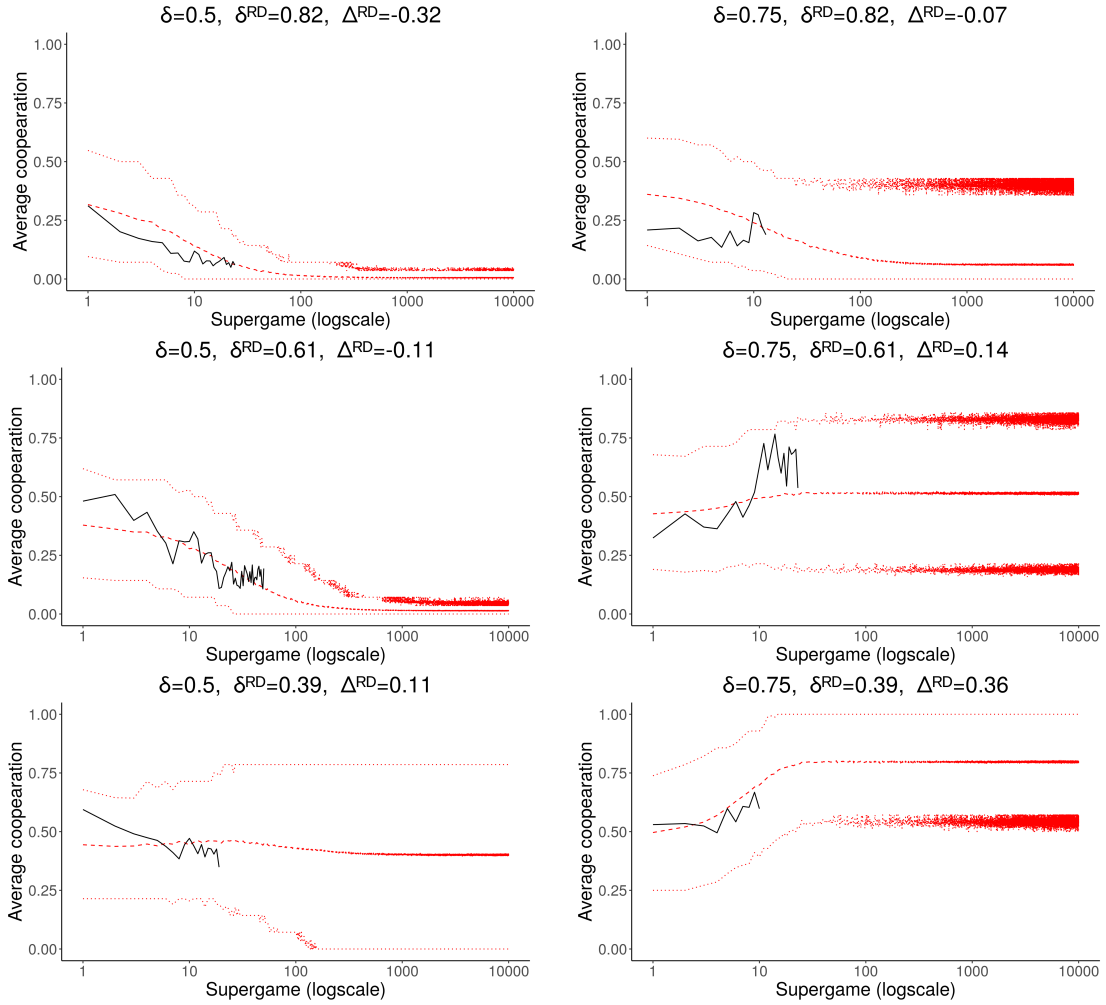

Figure 8: Predictions and actual behavior for six different treatments. The solid black line corresponds to the data, the red lines depict average cooperation and the middle 90% interval in 1,000 simulated populations.

Figure 8 displays the actual data and our confidence intervals for the time paths

of cooperation in these treatments.[16]

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 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 9 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.

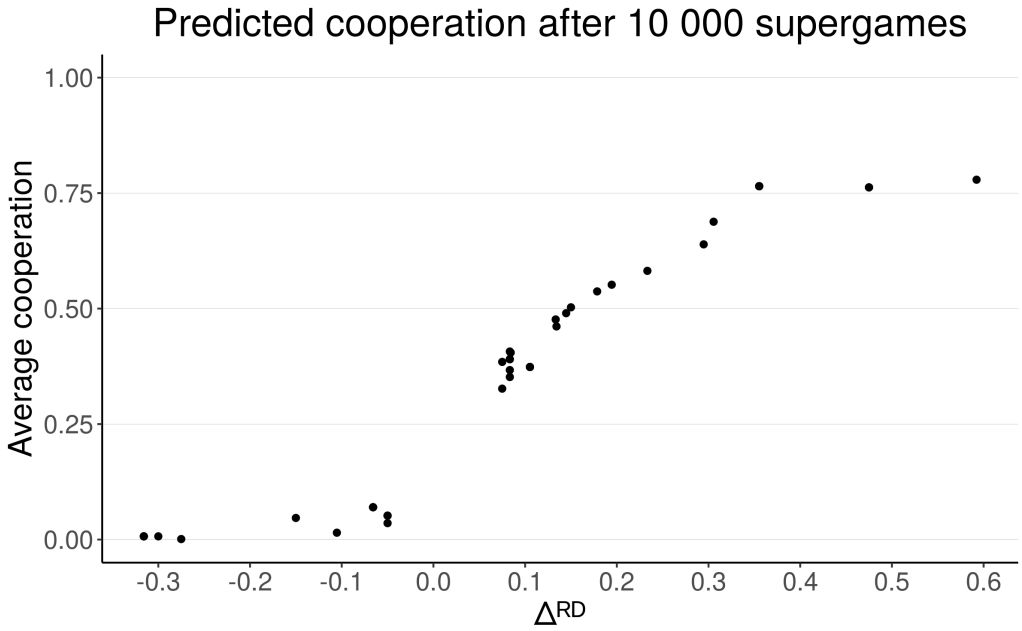

Figure 9: Predicted average cooperation after 10,000 supergames

[16]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.

7 Predicting the Next Action Played

Much of the existing literature has focused on determining the strategies used by the participants. Typically this is done with a mixture model, as in the SFEM of Dal Bó and Fréchette (2011), which assume that a finite number of strategies are used in the population. These models are often estimated using maximum likelihood, and their in-sample performance compared with information criteria. However, we want to compare the out-of-sample performance of our models with that of machine learning algorithms, so we instead consider how well the models predict the next action taken by a participant, given the participant's complete history so far. [17]

We consider mixtures of different numbers of types and also consider variants with a share of agents who always defect.[18] We compare the models to a naive benchmark that always predicts that $a_i(t+1) = a_i(t)$ and to a GBT.[19] As we show in Online Appendix E, the simple, six-parameter learning model predicts behavior almost as well as more complicated models. Furthermore, a learning model with the same mixture of types across treatments predicts the next action better than a pure strategy model with 11 different pure strategies estimated separately for each treatment.

The naive benchmark does very well: In 84.6% percent of the cases, the participants simply repeat the action they took in the previous round. If we assume the GBT captures most of the predictable regularity in the data, there is roughly an additional 6.3% of the observations that we can hope to capture with a better model. In contrast, the pure strategy model performs poorly.

The best-performing model has three types, with learning at all memory-1 histories. However, learning on its own captures most of the heterogeneity in behavior. A single 6-parameter learning with semi-grim type has an accuracy of 87.6%, while the

[17] The complete history includes all the actions taken by the individual and their partners in previous rounds of this supergame and all previous supergames, game parameters, realized supergame lengths so far, and realized payoffs.

[18] Previous studies have consistently found that a substantial share of participants behaves roughly this way, perhaps because they fail to understand the existence of repetitive equilibria.

[19] We remove the first round of the first supergame when calculating the performance of the naive predictions.

best performing model has 88.5%. Neglecting the influence of learning might lead researchers overemphasize the heterogeneity in participants when the diversity of play is largely driven by differences in experience.

8 Conclusion

This paper studies how to predict cooperation rates in the experimental play of the prisoner's dilemma as a function of the game parameters and the number of supergames played. We found that the key to predicting cooperation in a given match is the prediction of play in the initial round, and that this depends both on the game parameters and on the individual's experience in previous matches.

Our preferred learning model is very simple, as it holds play fixed except in the initial round of each supergame. This model only has 6 parameters and only one type of agent. While the 6-parameter model is too stark to model the richness of actual behavior, it predicts average cooperation at least as well as the more complex ML algorithms or more complicated learning models. This may be due in part be due to the lack of enough data, but also comes from the way our learning model uses data on the realized lengths of the supergames.

Our results lead to a clearer understanding of how and why the composite parameter Δ^{RD} influences cooperation rates: The parameter's main effect is 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.

Our model lets us capture the effect of playing more supergames on average cooperation. One advantage of this is that we can predict what average cooperation rates would be with longer lab sessions (assuming the participants did not loose focus on the task).

In this paper we only consider the prisoner's dilemma with perfect monitoring. Many real-world settings have implementation errors or imperfect monitoring and, as shown by Fudenberg, Rand and Dreber (2012), in such cases people seem to use more complex strategies with longer memory. There are not yet enough experimental studies of these games to support the sort of analysis we do here, but once there are it would be useful to extend our analysis of average cooperation rates to this case.

We close with a novel comparative statics prediction inspired by our learning model. In the lab, there is typically a tradeoff between specifying high discount factors and having participants play many supergames. So consider varying δ and g holding Δ^{RD} fixed, and suppose that the number of supergames played is inversely proportional to their expected length, which is $1/(1-\delta)$. Our model predicts that with more supergames and lower δ, there will be higher average cooperation if $\Delta^{RD} > .15$, and lower average cooperation if $\Delta^{RD} < 0$.

References

Aoyagi, Masaki, V. Bhaskar, and Guillaume R. Fréchette. 2019. "The impact of monitoring in infinitely repeated games: Perfect, public, and private." *American Economic Journal: Microeconomics*, 11: 1–43.

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

Backhaus, T., and Y. Breitmoser. 2018. "God does not play dice, but do we? On the determinism of choice in long-run interactions."

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: 164–192.

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."

Cheung, Y., and D. Friedman. 1997. "Individual Learning in Normal Form Games :." *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. R. Fréchette. 2011. "The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence." *American Economic Review*, 101: 411–429.

Dal Bó, P., and G. R. 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. R. Fréchette. 2019. "Strategy Choice in the Infinitely Repeated Prisoner's Dilemma." *American Economic Review*, 109: 3929–3952.

Engle-Warnick, J., and R. L. Slonim. 2006. "Learning to trust in indefinitely repeated games." *Games and Economic Behavior*, 54: 95–114.

Erev, I., and A. E. Roth. 1998. "Predicting how People Play Games."

Erev, I., and A. E. Roth. 2001. "Simple Reinforcement Learning Models and Reciprocation in the Prisoner's Dilemma Game." In *Bounded rationality: The adaptive toolbox.* , ed. Gerd Gigerenzer et al., Chapter 12. The MIT Press.

Fudenberg, D., and A. Liang. 2019. "Predicting and Understanding Initial Play." *American Economic Review*, 109: 4112–4141.

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

Fudenberg, D., J. Kleinberg, A. Liang, and S. Mullainathan. 2020. "Measuring the Completeness of Theories."

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

Harrington, Joseph E. 2017. *The Theory of Collusion and Competition Policy.* The MIT Press.

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

Honhon, Dorothée, and Kyle Hyndman. 2020. "Flexibility and reputation in repeated Prisoner's dilemma games." *Management Science*, 66: 4998–5014.

Ioannou, C. A., 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: 94–105.

Lundberg, S. M., and S. Lee. 2017. "A Unified Approach to Interpreting Model Predictions." 4765—-4774.

Mengel, Friederike, Simon Weidenholzer, and Ludovica Orlandi. 2021. "Match Length Realization and Cooperation in Indefinitely Repeated Games." *Available at SSRN 3777155*.

Mishra, S. K. 2016. "Journal of Economics." *Journal of Economics Bibliography*, 3: 498–515.

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. G., and M. A. Nowak. 2013. "Human cooperation." *Trends in Cognitive Sciences*, 17: 413–425.

Romero, J., and Y. Rosokha. 2018*a*. "Constructing strategies in the indefinitely repeated prisoner's dilemma game." *European Economic Review*, 104: 185–219.

Romero, J., and Y. Rosokha. 2018*b*. "Mixed Strategies in the Indefinitely Repeated Prisoner's Dilemma." *SSRN Electronic Journal*.

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

Wright, J. R., 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[20] 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 variables for all 30 candidates. The best of these parameters are then returned as the solution.

[20] From the package BlackBoxOptim.jl

B Complete Prediction Results

Table 9: Out-of-sample prediction loss (MSE) for per-session average cooperation

Model	Avg C MSE	S.E.	Improvement
Constant	0.0517	(0.0011)	-
OLS on Δ^{RD}	0.0189	(0.0006)	63.44%
OLS	0.0153	(0.0005)	70.40%
GBT	0.0151	(0.0005)	70.79%
Lasso	0.0145	(0.0005)	71.95%
Pure strategy belief learning w/o trembles	0.0436	(0.0010)	43.38%
Pure strategy reinf. learning w/ trembles	0.0396	(0.0008)	48.57%
Pure strategy belief learning w/ trembles	0.0379	(0.0008)	50.78%
Learning at all h two rates	0.0321	(0.0006)	58.31%
Learning with semi-grim and recency	0.0321	(0.0006)	58.31%
Learning at all h	0.0317	(0.0006)	58.83%
Learning with flexible memory-1	0.0313	(0.0006)	59.35%
Learning with memory-1	0.0311	(0.0006)	59.61%
Learning with semi-grim and AllD	0.0311	(0.0006)	59.61%
Learning with semi-grim	0.0310	(0.0006)	59.74%
Learning with semi-grim, two types	0.0304	(0.0006)	60.52%

Table 10: Out-of-sample prediction loss (MSE) for the Time Path of Cooperation

Model	Time-path MSE	S.E.	Improvement
Constant	0.0770	(0.0014)	0.0%
OLS on Δ^{RD}	0.0399	(0.0007)	48.18%
OLS	0.0325	(0.0005)	57.79%
Lasso	0.0325	(0.0005)	57.79%
GBT	0.0322	(0.0006)	58.18%
Pure strategy belief learning w/o trembles	0.0436	(0.0010)	43.38%
Pure strategy reinf. learning w/ trembles	0.0396	(0.0008)	48.57%
Pure strategy belief learning w/ trembles	0.0379	(0.0008)	50.78%
Learning at all h two rates	0.0321	(0.0006)	58.31%
Learning with semi-grim and recency	0.0321	(0.0006)	58.31%
Learning at all h	0.0317	(0.0006)	58.83%
Learning with flexible memory-1	0.0313	(0.0006)	59.35%
Learning with memory-1	0.0311	(0.0006)	59.61%
Learning with semi-grim and AllD	0.0311	(0.0006)	59.61%
Learning with semi-grim	0.0310	(0.0006)	59.74%
Learning with semi-grim, two types	0.0304	(0.0006)	60.52%

C Game parameters and learning

The 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

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

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.

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

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

To calculate the relative importance of Δ^{RD} we take the average of the variation introduced by Δ^{RD} alone, and the additional variation when it is added to the direct effect of $e_i(s)$ divided by the total variation.

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

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

This is the Shapley value of the two effects[21] It can be calculated on either the individual or treatment level, where the probabilities $p_i(s)$ are first averaged for each session.

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

Data	$Var(p\|e, \Delta^{RD})$	$Var(p\|e)$	$Var(p\|\Delta^{RD})$	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 11: Relative importance measures.

We see that 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, so in our model experience drives most of the variation initial round cooperation.

D Pairwise tests

Here we consider paired t-tests and paired signed Wilcox tests of whether the out-of-sample predictive MSE of the various models are significantly different. With the 10 different 10-fold cross-validation splits, we have 100 different test sets. Since we use the same splits for all predictive models, we can do paired tests. In the tables below paired tests between the initial round learning model and alternatives are shown for the average cooperation prediction task and for the time-path prediction task.

Table 12 shows that learning with semi-grim is significantly better than almost all alternatives, including most generalizations of the model, and the differences with the generalizations are small.

Model	Difference	T-test p-value	Sign test p-value
Pure strategy reinf. learning w/ trembles	-0.0036	p<0.001	p<0.001
OLS	-0.0014	p<0.001	p<0.001
GBT	-0.0013	p<0.001	p<0.001
Lasso	-0.0007	p=0.013	p=0.015
Learning with semi-grim and AllD	-0.0004	p=0.038	p=0.018
Learning with semi-grim and recency	-0.0004	p=0.013	p=0.031
Learning at all h two rates	-0.0002	p=0.229	p=0.168
Learning with flexible memory-1	-0.0002	p=0.293	p=0.473
Learning with memory-1	-0.0001	p=0.352	p=0.211
Learning at all h	-0.0	p=0.984	p=0.700
Learning with semi-grim, two types	0.0003	p=0.106	p=0.041

Table 12: Paired significance tests vs. learning with semi-grim for predicting average cooperation.

Model	Difference	T-test p-value	Sign test p-value
Pure strategy belief learning w/ trembles	-0.0069	p<0.001	p<0.001
OLS	-0.0015	p<0.001	p<0.001
Lasso	-0.0014	p<0.001	p=0.001
GBT	-0.0012	p=0.002	p=0.009
Learning at all h two rates	-0.0011	p<0.001	p<0.001
Learning with semi-grim and recency	-0.0011	p<0.001	p<0.001
Learning at all h	-0.0006	p=0.018	p=0.027
Learning with flexible memory-1	-0.0003	p=0.300	p=0.708
Learning with memory-1	-0.0001	p=0.533	p=0.645
Learning with semi-grim and AllD	-0.0001	p=0.629	p=0.767
Learning with semi-grim, two types	0.0006	p=0.017	p=0.004

Table 13: Differences and paired significance test with the main learning model (learning with semi-grim) for the time-path prediction task.

For predicting time paths, the picture is similar. The semi-grim learning model with two types is marginally better, and here it is significant at the 5% level.