Learning about Initial Play Determines Average Cooperation in Repeated Games*

Drew Fudenberg[1] and Gustav Karreskog[2]

[1]Department of Economics, MIT

[2]Department of Economics, Stockholm School of Economics

November 13, 2020

Abstract

We propose a simple learning model to predict cooperation rates across treatments in the experimental play of the indefinitely repeated prisoner's dilemma. Although the model has only 4 parameters, it performs almost as well as more complicated models and machine learning algorithms. We find that learning has the most effect on choices in the initial round of each supergame, and that whether cooperation rises or falls in the course of a session depends on the way the initial choices in a supergame determine play in subsequent rounds. Our results also explain past findings on the impact of the risk dominance considerations.

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

*We thank Anna Dreber Almenberg, Ying Gao, Annie Liang, Indira Puri, Emanuel Vespa, Jörgen Weibull, and David Rand for helpful comments, and NSF grant SES 1643517 for 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 as a guide for 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. Using data from the 32 treatments gathered in Dal Bó and Fréchette (2018), we show that average cooperation within a given supergame is mostly determined by the actions chosen in its initial round, and that the distribution of the initial actions changes over the course of an experimental session as the participants learn from feedback. Motivated by these observations, 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, and if they both defected they keep defecting. If they mismatch, i.e., one player cooperated and one defected, they both cooperate with roughly 1/3 probability (the overall average in our data.)

With this model, the way that people play in the first round of their very first supergame depends on a composite parameter Δ^{RD}, which measures 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 in a hypothetical game that we explain below. 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 both Δ^{RD} and on the initial actions and associated overall payoffs the player received in past supergames.

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 prelimimnary 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.

We estimate the parameters of the learning model in two ways: By finding the parameters that best predict average cooperation on the training sets, and by finding the parameters that best predict the time-path of cooperation. We find that estimating the model to fit the time paths yields lower cross-validated prediction error for the average cooperation rates, perhaps because there are too few sessions (103) to get reliable parameter estimates when treating each session as one observation. The learning model predicts the time-path of cooperation better than any of the black-box methods we consider, and is at least as good as them for predicting average cooperation in a session. Furthermore, we find that allowing for heterogeneous agents, or a more complex learning model with learning at all memory-1 histories, gives only modest improvements.

The learning model also allows us to predict what would happen with 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, both initial round cooperators and 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 of simply predicting that a player's current action will be the same as their previous one fits quite well. Moreover, our four-parameter learning model does as well at predicting behavior as a pure strategy model that incorporates 12 different pure strategies estimated separately for each treatment. Adding a second type that always plays D (and an accompanying fifth parameter for its share) leads to better predictions. At the cost of using more parameters and possibly making the model less robust, we can improve the model yet again by extending learning to all memory-1 histories or by allowing for multiple learning types.

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 does a good job at both overall predicting average cooperation and its time path.

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).

[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

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

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 below, we develop a very simple model that assumes all individuals use memory-1 strategies, and moreover

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

that these strategies differ across individuals only with respect to play in the initial round of each supergame. We assume that this initial behavior is driven by two different 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 $e_i(s)$. Specifically, we suppose that after each supergame s, $e_i(s)$ is updated based on the action $a_i(s) \in \{-1, 1\}$ taken, where -1 corresponds to D and 1 to C, and the total payoff received in supergame s, $V_i(s)$. We capture the evolution of the experience with

$$e_i(s) = \beta_{reinforce} \cdot a_i(s-1) \cdot V_i(s-1) + \rho \cdot e_i(s-1), \tag{1}$$

where $\beta_{reinforce}$ determines the strength of the learning, ρ discounts previous experiences, and we set $e_i(1) = 0$. The direct effect of Δ^{RD} and the experience $e_i(s)$ is then combined into a logistic function to keep the probability of cooperation between 0 and 1,

$$p_i^{initial}(s) = \frac{1}{1 + \exp\left(-(\alpha + \beta \cdot \Delta^{RD} + e_i(s))\right)}. \tag{2}$$

Cooperation or defection in the initial round is thus reinforced depending on the resulting supergame payoffs, while the direct influence of Δ^{RD} is constant across supergames.

Importantly, 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 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. We assume that $\sigma_{CC} = 0.960, \sigma_{DC} = 0.348, \sigma_{CD} = 0.315, \sigma_{DD} = 0.052$, which are the corresponding empirical frequencies in the data.

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 with high cooperation rates than the sign of $(\delta - \delta^{\text{SPE}})$.[3]

Several papers have attempted to 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.[4]

Recently, studies have found 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. We will use strategies similar to this semi-grim form

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

[4]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.

in our main learning model.[5]

Dal Bó (2005) and subsequent work has established 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 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. Erev and Roth (2001) considers a reinforcement model where TFT, and a lenient TFT, are added to always C and Always D as possible strategies. [6]

Two other established empirical regularities related to learning are that cooperation in the first round is increased if the realized length of the previous supergame is longer than expected, and if the partner cooperated in the first round of the previous supergame (Engle-Warnick and Slonim, 2006; Dal Bó and Fréchette, 2011, 2018). These two effects also point to a model of behavior where some form of reinforcement or learning drives cooperation in the initial round.

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.

[5]Romero and Rosokha (2018*b*) elicits memory-1 mixed strategies and finds that a finite 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 $\sigma_{CD} = .35, \sigma_{DC} = .32$.

[6]see Hanaki et al. (2005) and Ioannou and Romero (2014) study learning *within* very long supergames.

4 Summary of the data

We analyze the meta-data from Dal Bó and Fréchette (2018), who included experiments on the repeated prisoner's dilemma with perfect monitoring that were published before 2014. We consider only their treatments with $\delta > 0$. Our resulting data set contains observations from 12 different papers, 32 different treatments, and 103 incentivized experimental laboratory sessions, with 1,734 distinct participants and 145,802 individual choices. Here we highlight some aspects of the data that are of particular relevance to our work.

The discount factors ranged from 0.5 to 0.95. In 7 of the sessions, $\delta < \delta^{\text{SPE}}$, so no cooperation can occur in a subgame perfect equilibrium. In 19, cooperation can be supported by a SPE, i.e. $\delta > \delta^{\text{SPE}}$, but $\delta < \delta^{\text{RD}}$, so it is not risk dominant in the sense of Blonski, Ockenfels and Spagnolo (2011). In the remaining 77 sessions, $\delta > \delta^{\text{RD}}$.

The average rate of cooperation over all sessions was 41.3%. It was 9.3% for games where $\delta < \delta^{\text{SPE}}$, 18.7% for $\delta^{\text{SPE}} < \delta < \delta^{\text{RD}}$, and and 49.6% for $\delta > \delta^{\text{RD}}$.

The average play after the different memory-1 histories, and their frequencies, are shown in table 1. We see that the CD and DC histories are only a small subset of observations, roughly 15% together. Furthermore, we see that the average behavior is closed 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	96.0 %	34 395
DC	34.8 %	11 006
CD	31.5 %	11 006
DD	5.2 %	50 305
Initial	44.3 %	39 090
Total	41.3 %	145 802

To visualize how behavior differs depending on Δ^{RD} we group the sessions in the 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 a look at the data. The thresholds and relative frequencies of Δ^{RD} can be seen in figure 2.

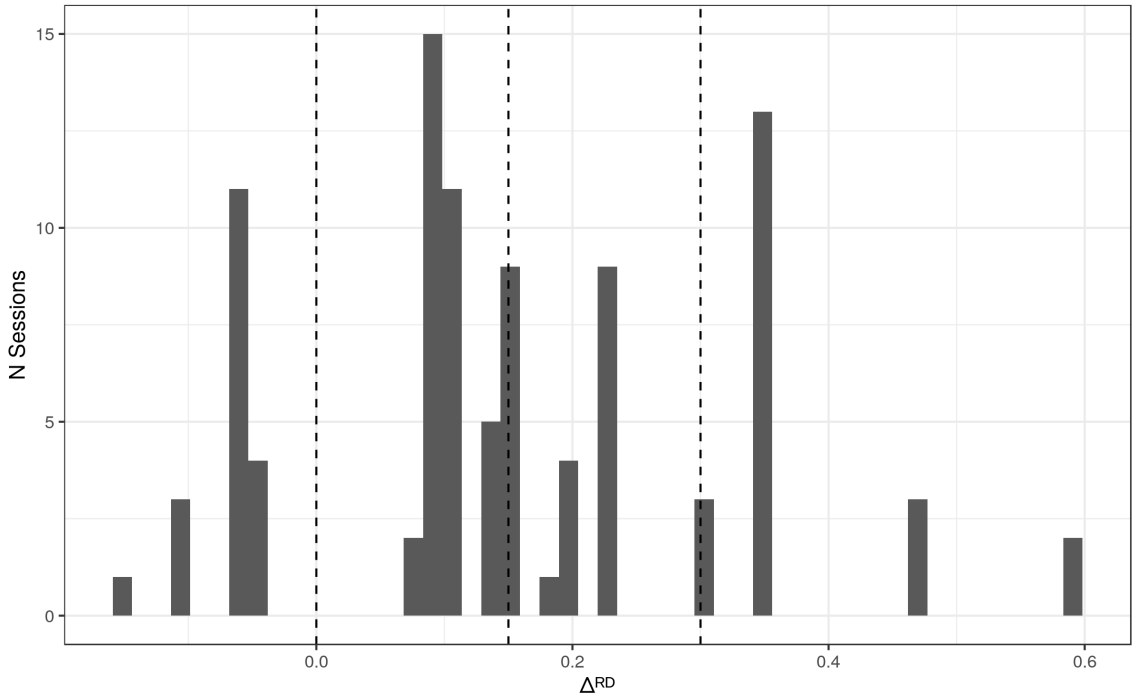

Figure 2: Distribution of Δ^{RD} in experimental sessions with $\delta > \delta^{SPE}$. The dashed vertical lines show to the thresholds for the groups.

Figure 3 shows the evolution of cooperation during the first 10 supergames, restricted to the sessions that played at least 10 supergames (86 of 103), and in figure 4 the first 20 supergames restricted to the sessions that included at least 20 supergames (50 of 103).

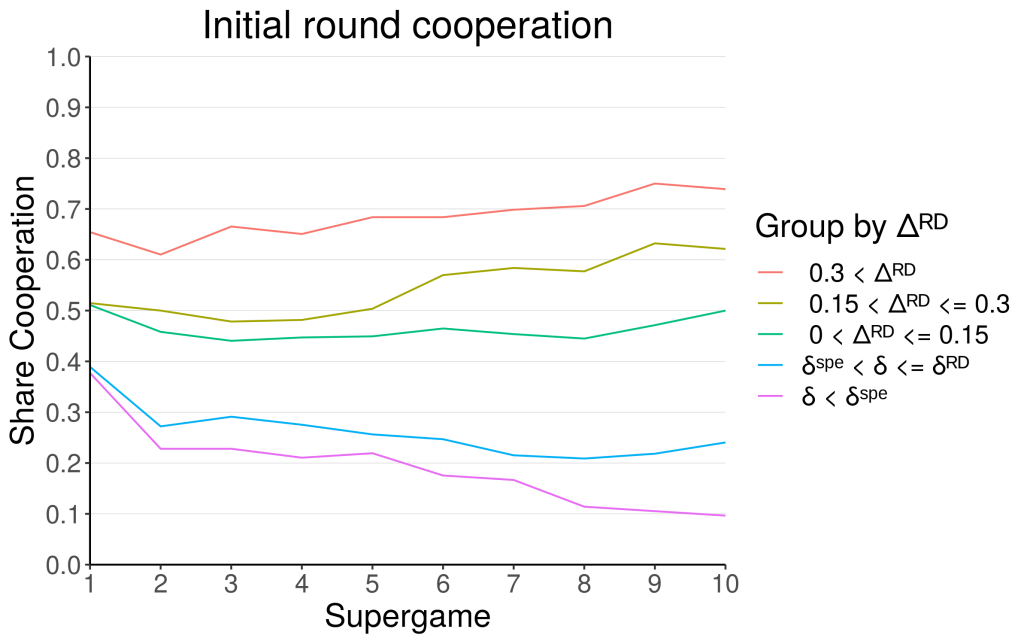

Figure 3: Average cooperation in the initial round over the 10 first supergames.

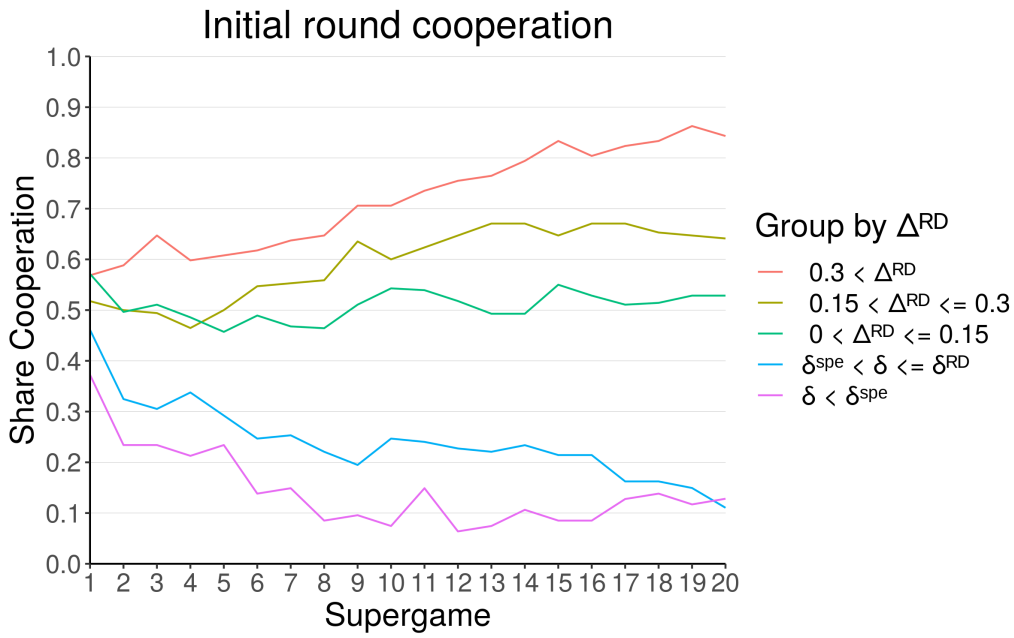

Figure 4: Average cooperation in the initial round over the 20 first supergames.

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 Appendix A this pattern is less sharp after the other memory-1 histories. This suggests that the differences in average cooperation across different 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 was played 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 9 of Appendix A, 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 the 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 improve those predictions. Similarly, we will find that the fit of our simple learning model is only slightly 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.[7] 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.

[7]Backhaus and Breitmoser (2018) suggests that the insensitivity of non-initial play to game parameters occurs because people ignore the game parameters and treat all repeated games the same, but this does not fit the fact that Δ^{RD} helps predict initial-round cooperation.

5 Black-Box Predictions of Play

Before we introduce the learning model and its predictive performance, we use ordinary least squares (OLS), Lasso, and gradient boosting trees (GBT) to predict the average level of cooperation and its time path in each of the 103 experimental sessions.[8] These estimates will serve as benchmarks for evaluating the performance of the learning model.

Throughout the paper, we evaluate models based on their estimated out of sample predictive performance. This lets us reduce problems with overfitting, and make sure that the regularities we find actually improve predictions. In general, out of sample predictions will favor simpler models that rely on stable predictors, something we see in this paper as well. 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. [9]

To estimate out of sample predictive performance we use a 10-fold cross-validation, which means that we divide the sessions into 10 different folds.[10] 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. Since play in a session is strongly correlated across supergames, we make the train/test splits on the level of the session, so each observation is predicted using only data from other sessions. To estimate the standard errors of the estimated mean squared error (MSE), we do 10 different such cross-validations, and use the variation between these 10 different cross-validations to

[8]The details of GBT vary. We used the implementation of LightGBM Ke et al. (2017). See the Online Appendix for more details. We also considered random forests, but do not report those results because they did not perform as well.

[9]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.

[10]see e.g. Hastie, Tibshirani and Friedman (2009) for an explanation of cross-validation.

estimate the standard errors of the out of sample MSE prediction error. [11]

As we will see, Δ^{RD} is a strong predictor of per-session average cooperation. While a larger set of features and more complicated predictive algorithms do improve predictions slightly, a linear function of Δ^{RD} captures most of the difference between a constant prediction benchmark and the best performing black-box predictions.

5.1 Predicting Average Cooperation

To make out of sample predictions of average cooperation, our 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 $\Delta^{RD} > 0.0$), and some interaction terms. In the online appendix we report in-sample regressions with three different selections from this feature set. The single regressor Δ^{RD} achieves an in sample MSE of .0178, while the full feature set has an in sample MSE of .0147. Table 10 in the appendix reports the out of sample MSE, and shows that a linear function of Δ^{RD} captures most of the possible improvement compared to the constant prediction.[12] In contrast to the in-sample MSE, adding more features to the OLS improves the MSE of out of sample predictions only slightly. The OLS with just Δ^{RD} performs better than the GBT, though not as well as Lasso. On a sufficiently large data set we would expect both of these algorithms to outperform OLS, but that did not seem to be the case here, perhaps because our data set is too small.

5.2 Predicting the Time-Path of Cooperation

By the time path of cooperation, we mean the average cooperation in each round of each supergame of an experimental session. Understanding this time path is interesting in its own right. It will also let us better predict how the average cooperation depends

[11] We caution that accurate estimates of the standard errors require more data than we have available, because each observation is used multiple times, so the different estimates are not independent. We consider alternative hypothesis tests in appendix F

[12] The full OLS and the Lasso model use the same variables as the in-sample regressions, and are reported in the online appendix.

on the number of supergames and the realized lengths of the supergames. Moreover, as we will see, getting the time path right has the added benefit of improving out of sample predictions of average cooperation.

To predict time paths, we start with the same features as for predicting average cooperation, and then add the realized length of the previous supergame, an indicator for the initial round, round number, and the supergame number, along with some interaction terms.

In the online appendix we report in-sample regressions of the time-path of cooperation. The single regressor Δ^{RD} achieves an in sample MSE of .0388, while the full feature set has an in sample MSE of .0333. Most of the parameter estimates are significant, and go in the expected direction, with g and l being significantly negative while δ is significantly positive. Appendix B reports the out of sample MSE for various algorithms. As with predicting average cooperation, a single regressor with Δ^{RD} captures much of the possible improvement over the constant prediction. Including the full set of regressors only reduces the MSE from 0.0395 to 0.0370; GBT has the best performance with a MSE of 0.0363. We see that out of sample the difference between using the single regressor Δ^{RD} and considering the full set of features is much smaller, which highlights the importance of using out of sample estimates.

6 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.

Our main learning model, presented above in equations (1) and (2), assumes all agents use the same learning rule, which is an over-simplification. 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 8, 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 8.3. 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 is the corresponding average in the data over all of the treatments.[13]

We relax the assumption of fixed behavior at non-initial histories in section 6.4, and we consider a more richly-parameterized model that lets play at these histories adjust based on feedback according to equation 1. This richer, more flexible model yields only a slight improvement in predictions.

6.1 Estimation Method

To generate the learning model's predictions for an experimental session ζ, we take as input the game parameters $\Gamma(\zeta)$ and the realized sequence of supergame lengths $S(\zeta)$. 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 $e_i(2)$ are updated according to (1), using the simulated values. The learning thus takes place between supergames. The 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 experience $e_i(s)$ after each supergame. Thus the sequence of supergame lengths, and the total number of rounds in the session, enter the predictions through this simulation procedure.

[13]This means that the cross-validations we do are strictly speaking not pure cross-validations, since these constants are calculated on the full data set. However, the variation in these averages is very small across folds, and using the corresponding averages in the training data has almost no impact on either model fit or estimated parameters. See appendix C in the online appendix.

Once we have simulated a population like this, 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. Given the predictions, the simulations let us compute the approximate MSE of predictions. Appendix C gives a detailed description of the numerical process. In Appendix G we evaluate this estimation procedure on data simulated under different assumptions about how people actually behave.

6.2 Results

We use the learning model to predict average cooperation in two different ways. In the first, we simply find the parameters that best predict the average cooperation in the training set, and use those parameters to predict average cooperation in the test set. In the other approach, we find the parameters that best predict the *time-path of cooperation* in the training set and use those to predict the average cooperation in the test sets.

When we predict the time-paths of cooperation, our simple learning model outperforms our ML algorithms, and does so by about as much as the best ML algorithm (which here is GBT) outperforms the simple 1-variable regression.

Model	MSE	SE	Relative improvement
Constant prediction	0.0705	(0.0002)	0
OLS on Δ^{RD}	0.0395	(0.0001)	44.0%
OLS	0.0370	(0.0002)	47.5%
Lasso	0.0369	(0.0002)	47.6%
GBT:time-path	0.0363	(0.0002)	48.5 %
Learning	0.0338	(0.0001)	52.1 %

Table 2: Out of sample MSE for predicting the time-path of cooperation.

In figure 5 we show the out of sample predictions and actual values of cooperation in the initial round of the first 10 supergames. Here we use the initial round to reduce the noise introduced by changing supergame lengths. To get the out of sample

predictions, we use a 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 does overestimate the level of cooperation for the highest values of Δ^{RD}.

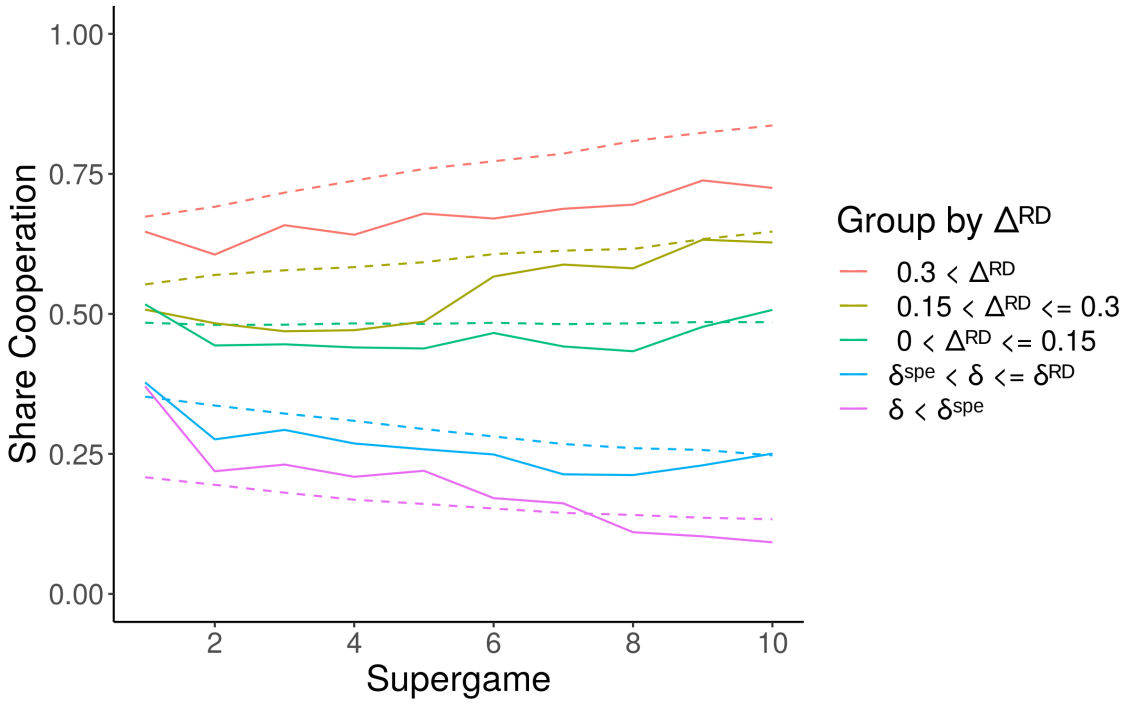

Figure 5: Actual (solid line) and out of sample predicted (dashed line) initial-round cooperation by supergame for all sessions that played at least 10 supergames grouped by Δ^{RD}.

The next table shows that estimating parameters from the time paths yields more accurate predictions of average cooperation than do the estimates based on average cooperation. It also shows that the learning model based on average cooperation does slightly worse than OLS of average cooperation on Δ^{RD}, while the learning model based on time paths performs as well as Lasso.

Table 3: Out of sample MSE for predicting average cooperation

Model	MSE	SE	Relative improvement
Constant prediction	0.0502	(0.0002)	0
OLS on Δ^{RD}	0.0184	(0.0001)	63.5%
OLS	0.0177	(0.0002)	64.7%
Lasso	0.0166	(0.0001)	66.9%
Learning estimated on avg	0.0173	(0.0002)	65.5%
Learning estimated on time path	0.0166	(0.0001)	66.9%

We might have expected to obtain the best prediction of average cooperation by using average cooperation to estimate the parameters, but we only have data on 103 sessions. As Table 4 shows, the standard deviation is much higher for the parameters estimated on average cooperation than those estimated on the time path, because here we use more of the data.

Variable	Estimated On	
	Time path	Average Cooperation
α	-0.38 (0.05)	-0.40 (0.13)
β	3.09 (0.32)	3.52 (0.46)
ρ	0.92 (0.03)	0.94 (0.13)
$\beta_{reinforce}$	0.13 (0.03)	0.40 (0.35)

Table 4: Average and standard deviations, in parentheses, for the parameter estimates across folds in the 10 different cross-validations.

From here on, we will focus on the parameters estimated on the time path, since those estimates are more stable and give better predictions. Remember that the experience is updated according to

$$e_i(s) = \beta_{reinforce} \cdot a_i(s-1) \cdot V_i(s-1) + \rho_i \cdot e_i(s-1).$$

which then enters into the probability of initial round cooperation by

$$p_i^{initial}(s) = \frac{1}{1 + \exp\left(-(\alpha + \beta \cdot \Delta^{RD} + e_i(s))\right)}. \tag{3}$$

The estimated $\alpha = -0.38$ means that for $\Delta^{RD} = 0$, about 40% of participants would cooperate before receiving any feedback. With $\Delta^{RD} = 0.1$ the probability of cooperation in the first supergame increases to 48.1%. The estimated recency effect is quite small, $\rho = 0.92$, which implies that the experience from a given supergame diminishes slowly. [14] In contrast, $\beta_{reinforce} = 0.13$ implies a strong learning effect: When the probability of cooperation in the initial round is .50, a positive feedback from cooperation of 1 will increase cooperation by .032. in the next supergame, and vice versa. This reinforcement can lead to a divergence in behavior, at least for a while, if both cooperation and defection give a positive payoff on average. As a consequence, in a given population the participants might behave quite differently even if they follow the same learning model.

As an example consider the case where $g = l = 2$, $\delta = 0.8$, and thus $\Delta^{RD} = 0$. If the first supergame an individual i plays goes the expected 5 rounds, and both partners cooperate all 5 rounds, the updated $e_i(2) = 0.13 \cdot 1 \cdot 5$, so i's probability of cooperation $p_i^{initial}(2)$ goes from 40.6% to 56.7%. An individual j experiencing DC in the first round and DD in the remaining 4 rounds gets payoff of 3, and so has $e_j(2) = 0.13 \cdot -3$, which implies that $p_j^{initial}(2)$ would go down to 31.6%.

Given these estimated parameters, the model can address the question of how much of the behavior is directly driven by the game parameters, and how much is driven by learning and so is only indirectly influenced by the game parameters. The two terms $(\alpha + \beta \cdot \Delta_i^{RD})$ and $e_i(s)$ enter into the equation (3) that determines initial round cooperation in the same way, so comparing how much of the variation in these two components drive variation in predicted behavior, we can get a sense of which effect is the more important factor for behavior. As we show in In appendix D, the resulting decomposition suggests that in the last supergame of each session, approximately 85% of the variation in individual behavior and 70% of the variation between sessions is driven by learning and not the direct influence of the game parameters. We also we show that individual behavior in the last supergames of the

[14] In fact, by setting $\rho = 1$ we don't increase MSE by a lot. However, especially when considering longer sessions, we think some recency effect is to be expected and we therefore include it in our main model.

session is consistent with those estimates.

6.3 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 initial round. 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 the simple learning model, we relate the supergame payoffs participants receive with 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 $\pi(C) - \pi(D)$ and Δ^{RD} in the data.

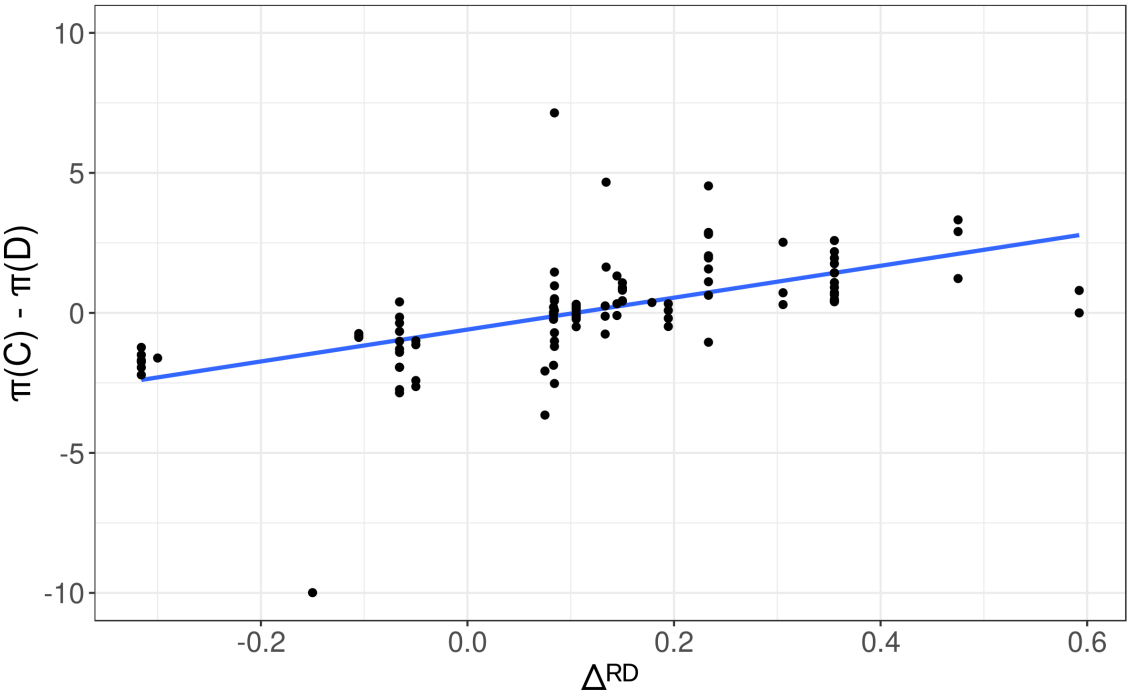

Figure 6: Average empirical difference between total payoff in supergames where the participant cooperated and defected. Each dot corresponds to one experimental session.

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. This helps explain why we see no clear trends in the sessions where $0 < \Delta^{RD} < 0.15$, cooperating and defecting are on average equally reinforced.

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 than actual participants, but it follows the same pattern as in the actual data.

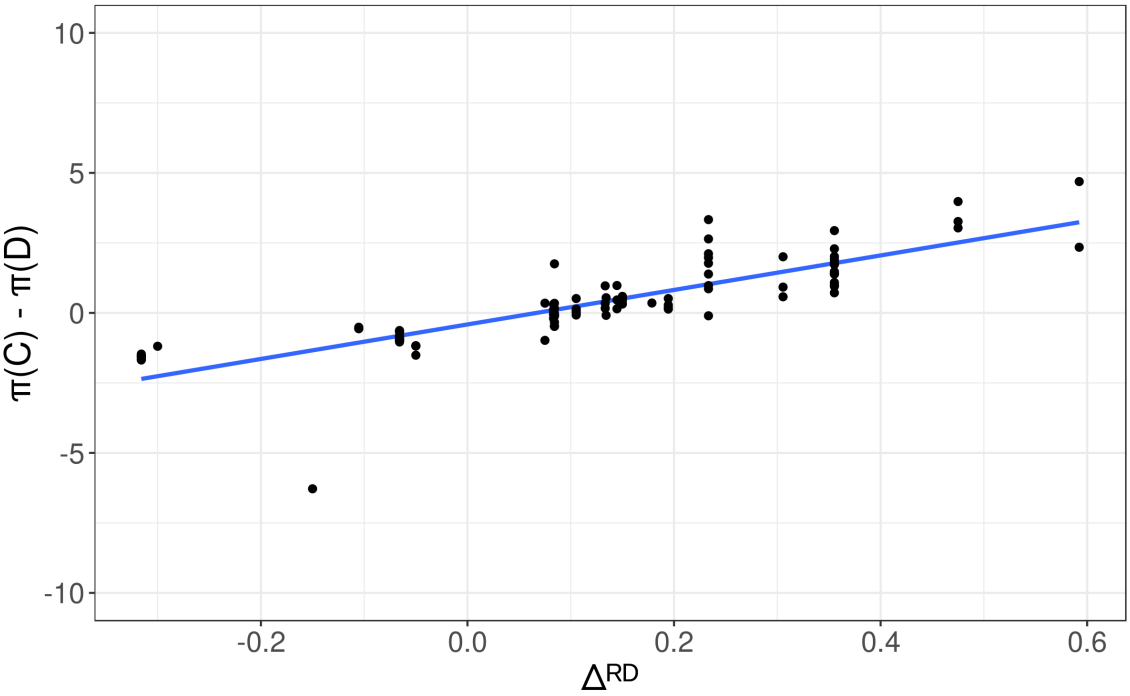

Figure 7: Average simulated difference between total payoff in supergames where the participant cooperated and defected. Each dot corresponds to one experimental session.

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 it capture some additional forces that determine cooperation, such as how many supergames were played and their realized lengths.

6.4 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 three questions: Can we improve our model by adding more parameters or by introducing heterogeneity? Would a different learning model perform as well?

Extending learning to 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 same learning dynamic we have for the initial round to all memory-1 histories. For each memory-1 history h, we track the experience at that history $e_i(h,t)$, where t now is a time variable running over all rounds and all supergames. Experience at h is only updated when h occurs, and when it does it depends on the individual's payoff for the rest of the supergame $V_i(t)$ according to

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

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

Of course, at time t, the individual does not know what $V_i(t)$ will turn out to be. Instead, the probabilities of cooperation are only updated in 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 learning model extends the number of parameters from 4 to 12, but does improve predictions slightly. Table 5 shows a comparison with our main learning model and all the different variations considered in this subsection.

Heterogeneous agents. It is commonly found that there is a lot of heterogeneity in individual behavior. To allow for this, we now consider a finite mixture extension of our learning model. We assume that there are two different types, with different parameters $(\alpha, \beta, \beta_{reinforce}, \rho)$, 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 only a slight difference.

When we consider the individual one-step ahead predictions, we will see that introducing heterogeneity does slightly improve prediction of both average cooperation

and the time-path of cooperation. One reason that the improvement is only slight may be that the learning model with a single type has endogenous heterogeneity in play that can account 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. [15] In contrast, adding a constant share of AllD players, if anything slightly decreased the accuracy of out of sample predictions.[16]

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 uses them to calculate the expected values from playing TFT or AllD. Given these 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 treatments prediction, increasing the original 6 parameters to 8. A more complete description of the model can be found in appendix E.

6.4.1 Results

In table 5 we see a comparison with the different alternatives considered in this subsection. We see that allowing learning to change play at each memory-1 history, and not only the initial one, slightly improves predictions, but at the cost of a more complicated model. The pure learning model does considerably worse, both with and without the introduction of noise, and on both the time path and the average cooperation. Lastly, including heterogeneity does not improve out of sample predictions on either prediction task.

[15] 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.

[16] 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.

Model	Time path	Average Cooperation
Constant prediction	0.0705 (0.0002)	0.0502 (0.0002)
OLS on Δ^{RD}	0.0395 (0.0001)	0.0184 (0.0001)
Lasso	0.0369 (0.0002)	0.0166 (0.0001)
GBT:time-path	0.0363 (0.0002)	
GBT:avgerage cooperation		0.0191 (0.0002)
Pure strategy belief learning	0.0419 (0.0003)	0.0211 (0.0001)
Pure strategy belief learning with noise	0.0387 (0.0002)	0.0207 (0.0001)
Initial round learning $\rho = 1$	0.0341 (0.0002)	0.0167 (0.0002)
Initial round learning	0.0338 (0.0001)	0.0166 (0.0001)
Initial round learning model with two types	0.0334 (0.0002)	0.0162 (0.0001)
Initial round learning and AllD	0.0347 (0.0002)	0.0171 (0.0002)
Learning at all memory-1 histories	0.0314 (0.0002)	0.0160 (0.0002)

Table 5: Out of sample MSE of predictions for time path and average cooperation for different learning models.

7 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.

7.1 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, but use only the first half of the sessions in each training set to estimate the parameters. We then 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.

Since the parameters estimated on the time path performed better in predicting

overall average cooperation, 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.

In table 6, we see the cross-validated MSE where we predict the later half of experimental sessions with parameters estimated on different models.

Model	Second half Session avg	Second half time path
Constant Prediction	0.071 (0.0002)	0.084 (0.0003)
OLS	0.032 (0.0003)	0.049 (0.0005)
Lasso	0.030 (0.0002)	0.044 (0.0005)
GBT:time-path	0.031 (0.0002)	0.043 (0.0002)
Initial round learning	0.027 (0.0003)	0.040 (0.0004)

Table 6: MSE from estimating on first half and evaluating on second half of the experimental sessions.

The table shows that the initial round learning model is better at extrapolating to longer supergames than our atheoretical black-box algorithms. While atheoretical prediction algorithms often perform very well withing a given domain, extrapolating to a slightly different settings is often difficult. As we see here, a simpler model that captures the essence of what is going can better extrapolate to related prediction problems.

7.2 Extrapolating to hypothetical session lengths

We now look at what would happen in the long run for the six different treatments in Dal Bó and Fréchette (2011). They encompass three different versions of the stage game, with two different values of δ. For each treatment, three sessions were run with on average 14 individuals per session.

We generate predictions by simulating the learning model with the average (across folds) parameters estimated on the time path in table 4. For each of the treatments, 1000 populations with 14 participants were simulated for 1000 supergames, with randomly drawn supergame lengths. Using these simulations we can compute both

the average simulated cooperation, and the 90% interval of the simulated average cooperation.

Even though we don't include ex-ante heterogeneity in this model, we still see quite wide 90% intervals in figure 8. The randomness in both behavior and the differing experiences can lead to substantially different outcomes. In the treatment $\Delta^{RD} = 0.11$, even after 1000 supergames the 90% interval goes from 28% to 100%, and the average is just 60%. For $\Delta^{RD} < 0$, we predict and see less than 50 % cooperation. For $\Delta^{RD} = 0.14$ we see a very slow increase in initial round cooperation, which hasn't yet stabilized after 1000 supergames and for $\Delta^{RD} = 0.31$ we see a relatively fast and certain prediction of high rates of cooperation.

Dal Bó and Fréchette (2011) estimate the learning model in subsection 6.4 on the individual level, and use those individual estimates to simulate behavior. They produce plots similar to the one below, but restricted to the initial round. Visual inspection suggests that our single model fits the data about as well, even though it is simpler and is designed to do prediction rather than in-sample replication.

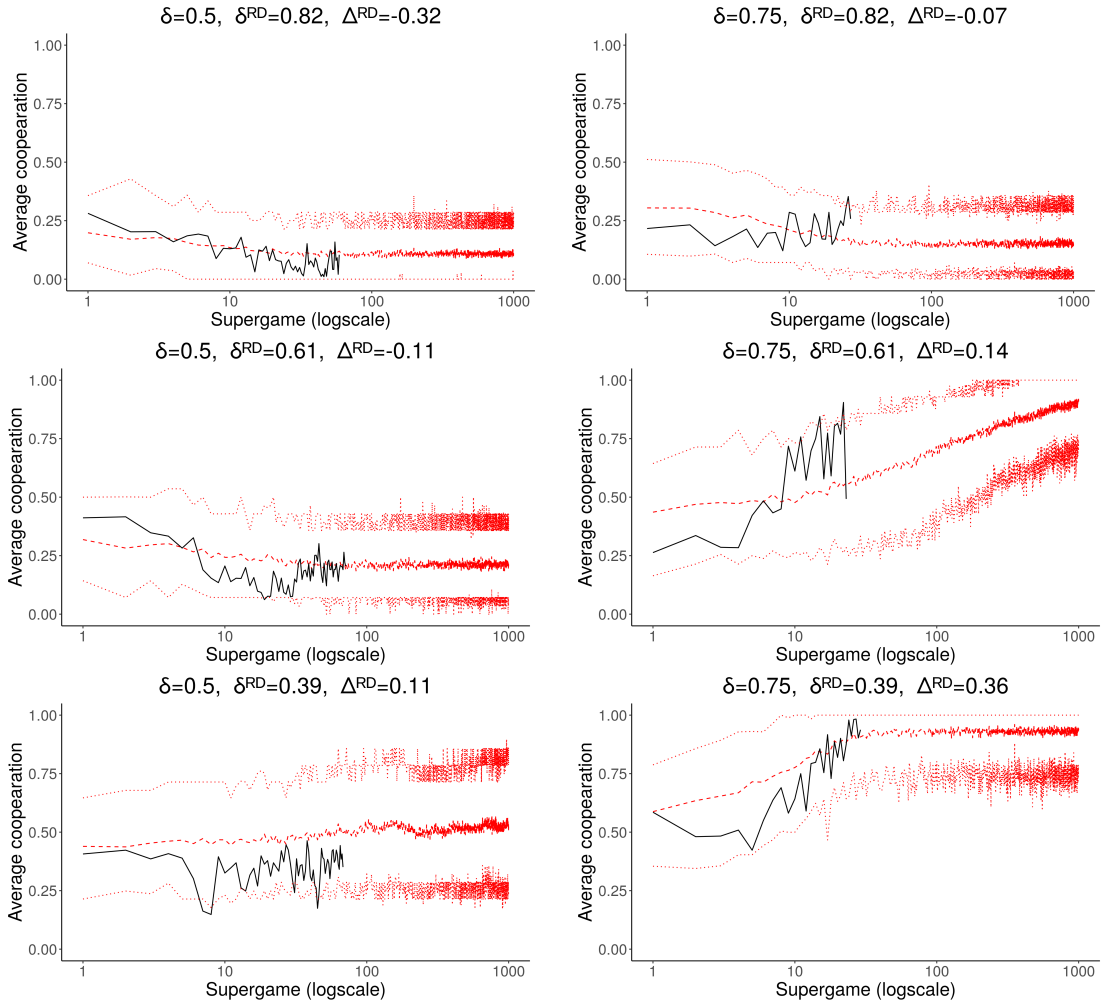

Figure 8: Long-run predictions and actual behavior for six different treatments. The solid black line corresponds to the average actual cooperation. The dashed red line is the average of 1000 simulated populations, the dotted red lines depict the middle 90% interval.

We get a broader picture of the long run predictions by replicating this exercise for all the 32 treatments in the data. In figure 9 we see the average cooperation after 10 000 supergames, predicted by simulating 1000 populations of size 16 for each treatment. We see that for Δ^{RD} between 0.1 and 0.3, even after 10 000 supergames, the initial round learning model does not predict either very high or very low rates of

cooperation.

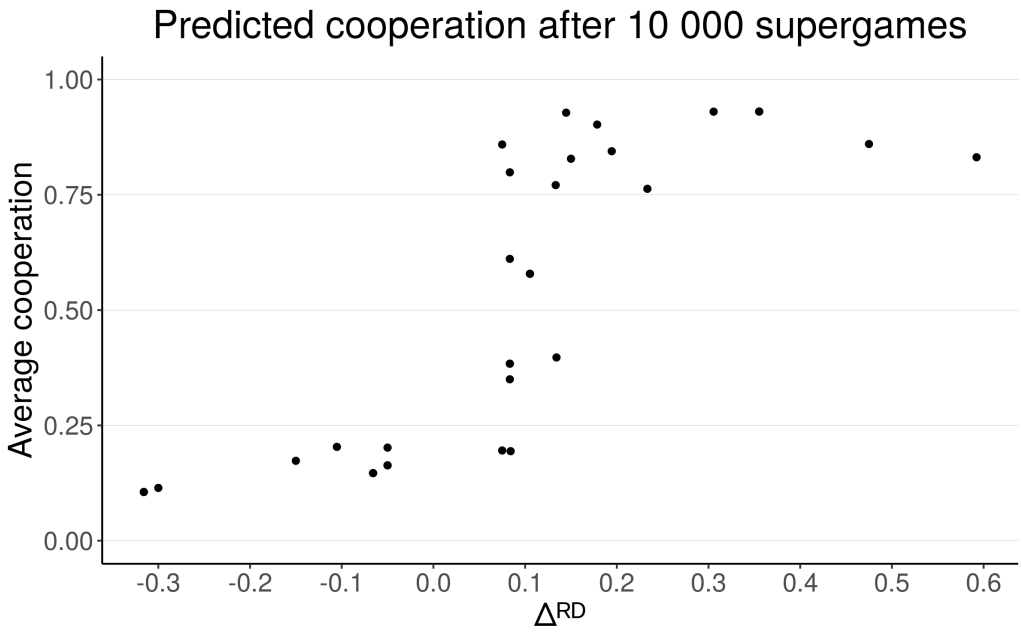

Figure 9: Predicted average cooperation after 10 000 supergames

For the two treatments with the highest value of Δ^{RD}, we see that the average cooperation goes down. This because these treatments have very high δ, and cooperation goes down slowly over time with the fixed mixed strategy we assume. If we instead look at initial round cooperation, we don't see this pattern anymore.

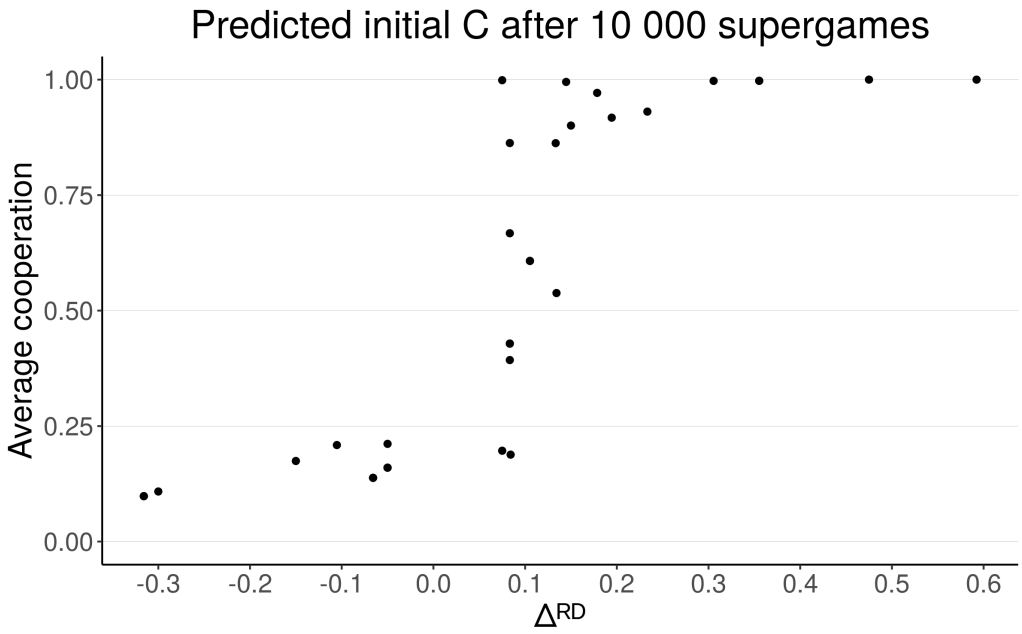

Figure 10: Predicted initial cooperation after 10 000 supergames

Furthermore, if we consider the full learning model, average cooperation does not decline for high values of Δ^{RD} because cooperation at all histories goes up. Moreover, in this very long run we see a clearer convergence to either cooperation or defection in the populations. However, this is only for the long run. After a 100 supergames, intermediate values of Δ^{RD} have intermediate rates of cooperation.

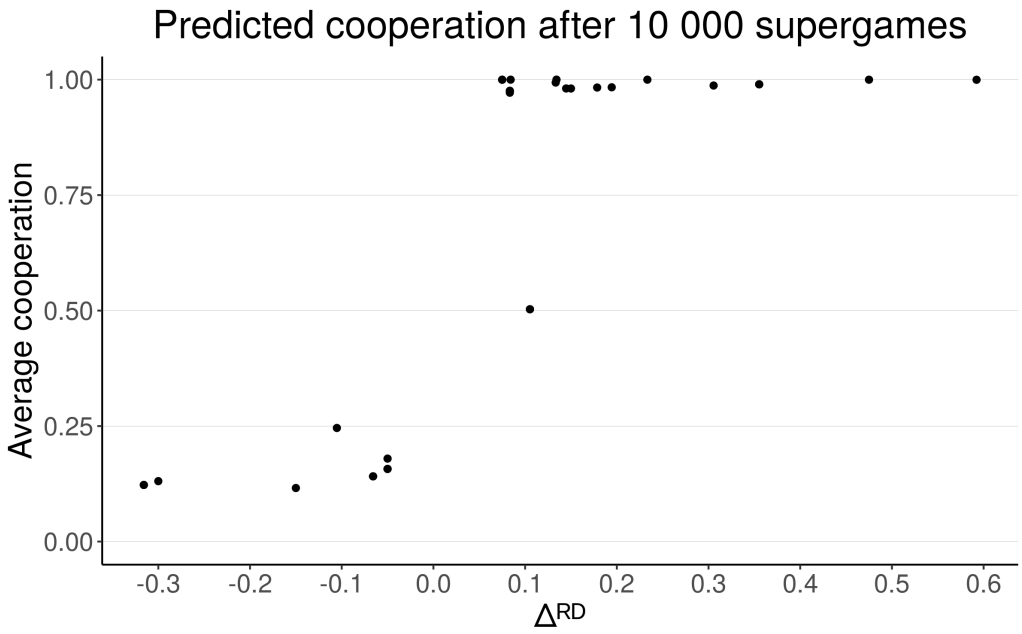

Figure 11: Predicted average cooperation after 10 000 supergames with the estimated full learning model.

8 Predicting the Next Action Played

A central question in the existing literature is to identify the strategies used by the participants. Typically this is done with a finite mixture model, as in the SFEM of Dal Bó and Fréchette (2011): Assume that a finite number of strategies are used in the population, and estimate the parameters and the shares of these strategies by maximum likelihood. The use of the maximum likelihood criterion makes it difficult to compare models of different complexities. We therefore consider the closely related problem of predicting the next action taken by a participant.[17] As we will see, the four-parameter model with initial round learning does as well at predicting behavior as a pure strategy model incorporating 12 different pure strategies estimated separately for each treatment. Adding a second type that always plays D (and accompanying

[17]In appendix H we report standard loglikehoods of a maximum likelihood estimations, which are consistent with the results of this main analysis.

parameters for its share and mistake probability) leads to better predictions. We can improve predictions even more by extending learning to all memory-1 histories or by allowing for multiple learning rules, but a single type first round learner together with a constant share AllD captures a substantial share of the predictable regularity.

8.1 Estimating and evaluating the models

We consider several different possible models of behavior: pure strategies with a fixed mistake probability, memory-1 mixed strategies, learning only in the initial round, and learning at all memory-1 histories.

The memory-1 mixed strategy model supposes that the probability of cooperation at each memory-1 history is given by a logistic function of Δ^{RD}, essentially the full learning model without any learning. In the pure strategy model we consider 12 different pure strategies, taken from the pure strategies estimated to have positive share in Fudenberg, Rand and Dreber (2012).[18] We allow the mistake probability to vary between the pure strategies, and estimate the shares of the different pure strategies and these mistake probabilities for each treatment separately.

The two learning models are the same as before. Since these models capture how participants adjust to the game parameters, we estimate them simultaneously on all treatments. In the 1-memory mixed strategy model the probability to cooperate after each history is given by a logistic function of Δ^{RD}. So the mixed strategy model is essentially the full learning model with the learning parameters set to 0.

To estimate out of sample predictive performance, we perform a 10-fold cross validation on the individual level. The models are all estimated as finite mixtures using maximum likelihood on the test data of each cross-validation split.

Let $D = \{(h_i(t), a_i(t))|i \in I, t \in T(i)\}$, be the collection of all individual histories and actions, and $m(h_i(t)) = \hat{a}_i(t)$ be model m's predicted probability of cooperation. Once estimated, a model m consists of a set of types σ^j, and corresponding shares in the mixture ϕ^j. Given those, we can calculate the probability of history $h_i(t)$

[18] While Fudenberg, Rand and Dreber (2012) studies interactions with exogenous noise, these 12 strategies contain those strategies often found to be used in games without noise e.g. Dal Bó and Fréchette (2018)

conditional on the different types in the finite mixture by

$$\Pr(h_i(t)|\sigma^j) = \prod_{\tau<t} \sigma^j(h_i(t))^{\mathbb{1}\{a_i(t)=1\}} \cdot \Big(1-\sigma^j(h_i(t))\Big)^{\mathbb{1}\{a_i(t)=-1\}}.$$

and the via Bayes's rule the probability of i being of type j

$$\Pr(\sigma^j|h_i(t)) = \frac{\phi^j \Pr(h_i(t)|\sigma^j)}{\sum_l \phi^l \Pr(h_i(t)|\sigma^l)}.$$

We then make the prediction $m(h_i(t)) = \sum_j \sigma^j(h_i(t))\Pr(\sigma^j|h_i(t))$.

We consider two different measures of predictive performance. The *prediction loss* is the average cross-entropy

$$\mathcal{L}(m|D',\theta) = \frac{-1}{|D'|} \sum_{(h_i(t),a_i(t))\in D'} \log(m(h_i(t)|\theta))\cdot\mathbb{1}\{a_i(t)=1\} + \log(1-m(h_i(t)|\theta))\cdot\mathbb{1}\{a_i(t)=-1\}.$$

The *accuracy* is the fraction of the predictions where the action taken was one that was predicted to be most likely. The *relative accuracy*

$$\frac{Accuracy(m|D,K) - Accuracy(m^{naive}|D,K)}{1 - Accuracy(m^{naive}|D,K)},$$

measures a model's improvement over the naive benchmark. If model m performs no better than the naive benchmark, then the relative accuracy is 0, and if the model m has an accuracy of 1 it also has a relative accuracy of 1.

8.2 Model Comparisons

The next table compares the performance of the initial-round and full learning models with the pure strategy model of Dal Bó and Fréchette (2011) and the memory-1 mixed strategy model. We consider different numbers of types in the finite mixtures, and we also considered variants with a share of agents who always defect.[19] We compare the models to a naive benchmark that always predicts that $a_i(t+1) = a_i(t)$ and to a

[19]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.

GBT.[20] The GBT also allows us to provide additional evidence for the assumption that memory-1 is sufficient to capture behavior. The online appendix compares the performance of some additional models.

Table 7: Out of sample one-step ahead predictive performance comparisons.

Model	N	AllD	Loss	Accuracy	Relative Accuracy
Naive			0.439	84.1%	
Pure			0.335	87.0%	18.6%
Memory-1 mixed	1		0.362	83.6%	-2.6%
	3		0.300	87.2%	19.6%
Initial round learning	1		0.328	87.2%	19.8%
	1	Yes	0.309	87.5%	21.4%
	3	Yes	0.297	87.9%	24.0%
Full learning	1		0.323	87.8%	23.2%
	3		0.285	88.3%	26.7%
GBT:next action			0.242	90.0%	38.6%

We can draw several conclusions from this exercise. Firstly, the naive benchmark is doing surprisingly well, in 84.1% 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% of the observations that we can hope to capture with a better model. In contrast, the pure strategy model performs poorly, both in terms of accuracy and prediction loss.

The best model is also the most complicated, three learning types and learning in all memory-1 histories. However, we also see that learning itself can capture most of the heterogeneity in behavior. A single mixed strategy model is a very bad description of individual behavior, even worse than our naive benchmark, but performance improves considerably if we allow for multiple types. In the case of the learning models, a single type already captures most of the behavior, and adding

[20]We remove the first round of the first supergame we calculating the predictive performance of the naive predictions.

multiple types give modest improvements.

Some of the remaining heterogeneity can be captured by introducing a share of AllD players, especially in terms of prediction loss. The six parameter model, Initial round learning and AllD, has an accuracy of 87.5%, which is actually slightly higher than the 32-parameter three type flexible mixed strategy model and only 0.8% less than the best performing 38-parameter full learning model. Neglecting the influence of learning might lead researchers overemphasize the heterogeneity in "types" when it is in fact to a large degree captured by differences in experiences.

A more complete version of Table 7 can be found in the online appendix. In that table we consider additional variations of the models presented above and also a hybrid model with initial round learning and flexible memory-1 mixed subsequent behavior. The performance of this hybrid model is in between that of simple initial round learning and the full learning model.

8.3 Gradient Boosting Trees and the memory-1 assumption

The previous literature typically finds that a small set of memory-1 pure strategies represent most of the behavior. To further evaluate this conclusion, table 8 compares three different GBT predictions, where the difference is the length of memory within each supergame. Each GBT has the summary statistics from previous supergame, but the memory-1 GBT only sees the outcome of the previous round of their current supergame while the memory-2 GBT sees the outcome of the previous two rounds. As we see, there is almost no difference in either the prediction loss or the accuracy.

Memory Length	Prediction loss	Accuracy
Memory-1 GBT:next action	0.244	89.97%
Memory-2 GBT:next action	0.243	90.01%
Memory-3 GBT:next action	0.242	90.05%

Table 8: Out of sample predictive performance for GBT with different within supergame memory-lengths.

This doesn't necessarily imply that participants are not using strategies with

longer memories. Since we only observe on-path behavior, it is still possible that they use strategies that use longer memories, but we don't observe these outcomes enough to detect the difference.

9 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 the 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 4 parameters, and one type of agent. We prefer the simpler model even though richer ones have lower cross-validated prediction errors due to our sense that the simpler model will have more external validity: While cross-validation helps guard against overfitting on our data, it is not clear how well its conclusions would extend to data drawn from a very different pool of subjects.

While the 4-parameter model is too stark to model the richness of actual behavior, it does as well as or better than more complex machine learning algorithms at predicting average cooperation rates. This may be due at least in part to the fact that our 103 sessions is a relatively small number of data points by machine learning standards; we expect that the ML algorithms would outperform our model given a sufficiently large data set. This is consistent with the fact that the GBT outperforms the learning model at predicting the next action played.

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 an out of sample 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 proportional to $1/(1-\delta)$. 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

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.

Ioannou, C. A., and J. Romero. 2014. "A generalized approach to belief learning in repeated games." *Games and Economic Behavior*, 87: 178–203.

Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu. 2017. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." In *Advances in Neural Information Processing Systems 30.* , ed. I. Guyon et al. Curran Associates, Inc.

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.

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 Additional Depictions of the Data

In figure 3 and 4 we saw that initial round behavior differs between different values of Δ^{RD}, and that the differences increase over time as the participants play more supergames. In figure 12 and 13 we show the corresponding plots but for different memory-1 histories.

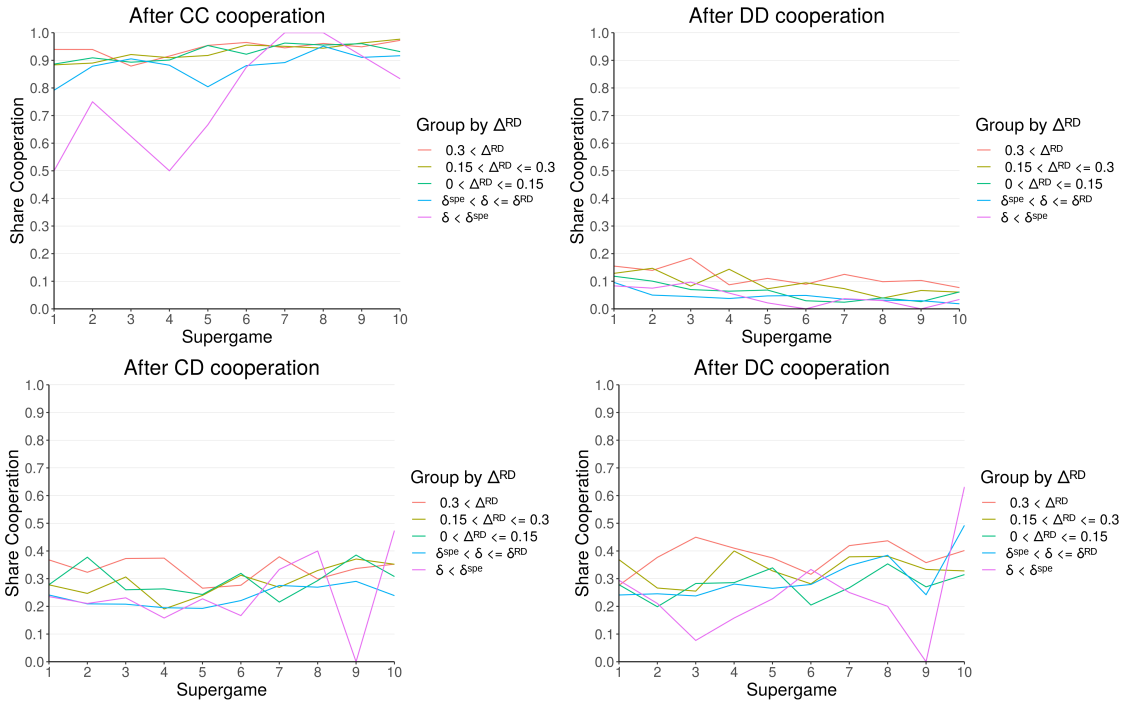

Figure 12: Average cooperation after different memory-1 histories for the first 10 supergames, grouped by Δ^{RD}.

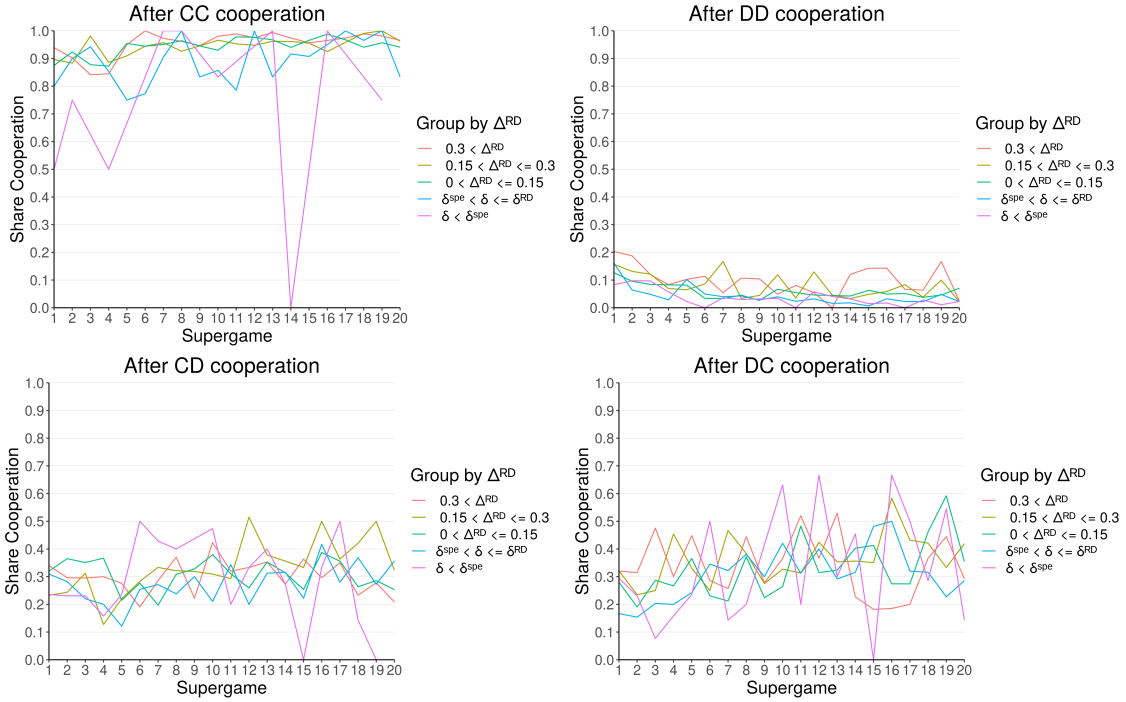

Figure 13: Average cooperation after different memory-1 histories for the first 20 supergames, grouped by Δ^{RD}.

Behavior at these memory-1 histories is less variable both between different values of Δ^{RD} and over the course of the experimental sessions. However, in contrast to the initial round, different players face different distributions of the other memory-1 histories, and because these differences are not exogenous there may selection effects. As a consequence, these plots should be interpreted with more caution. Furthermore, some memory-1 histories are uncommon in certain treatments, e.g. there are few CC in games where the average cooperation rate is low.

If the differences in average cooperation between different treatments is driven primarily by the initial round behavior, then average cooperation after the initial round should be should be primarily determined by the outcome of the initial round, and otherwise similar across treatments.

To show this we compare the following three regressions. The outcome variable is the average cooperation by a participant in a supergame in the rounds following the initial round, e.g., if 4 rounds were played in that particular supergame, we calculate the average cooperation by that participant in rounds 2, 3 and 4. The first regression, conditions only on the outcome of the initial round. The second adds

game parameters (δ, g, l, and Δ^{RD}), and the last uses only the game parameters and not the initial round.

Table 9: Rest of supergame average cooperation conditional on initial round outcome.

	(1)	(2)	(3)
initial = CD	-0.597^{***} (0.006)	-0.578^{***} (0.006)	
initial = DC	-0.604^{***} (0.006)	-0.585^{***} (0.006)	
initial = DD	-0.809^{***} (0.005)	-0.763^{***} (0.006)	
g		0.021^{***} (0.005)	0.015^{**} (0.007)
l		-0.009^{***} (0.003)	-0.034^{***} (0.004)
δ		-0.082^{**} (0.037)	-0.111^{**} (0.050)
Δ^{RD}		0.254^{***} (0.037)	0.944^{***} (0.049)
Constant	0.886^{***} (0.004)	0.884^{***} (0.017)	0.395^{***} (0.022)
Observations	25,574	25,574	25,574
R^2	0.533	0.537	0.170
Adjusted R^2	0.533	0.537	0.170
Note:			*p<0.1; **p<0.05; ***p<0.01

All the game parameters are statistically significant, but they explain almost no extra variance: The difference in R^2 between the model with and without game parameters is less than 0.01. In contrast, removing the outcome of the initial round lowers the R^2 to 0.0170. Table 3 in the Online Appendix shows that this is also true for second- round cooperation instead of average cooperation in the rest of the supergame.

B Black box prediction results

B.1 Average cooperation

Table 10: Out of sample MSE for per-session average cooperation

Model	MSE	SE	Relative improvment
Constant prediction	0.0502	(0.0002)	0
$\delta > \delta^{SPE}$	0.0432	(0.0002)	13.9%
$\delta > \delta^{RD}$	0.0291	(0.0001)	42.0%
OLS on (g, l, δ)	0.0239	(0.0003)	52.4%
GBT:average cooperation	0.0191	(0.0002)	62.0%
OLS on Δ^{RD}	0.0184	(0.0001)	63.5%
OLS-full	0.0177	(0.0001)	64.7%
Lasso	0.0166	(0.0001)	66.9%

B.2 Time path

Table 11: Out of Sample Predictive MSE for the Time Path of Cooperation

Model	MSE	SE	Relative improvement
Constant prediction	0.0705	(0.0002)	0
$\delta > \delta^{SPE}$	0.0631	(0.0002)	10.5%
$\delta > \delta^{RD}$	0.0490	(0.0001)	30.5%
OLS on (g, l, δ)	0.0423	(0.0002)	40.0%
OLS on Δ^{RD}	0.0395	(0.0001)	44.0%
OLS-full	0.0370	(0.0002)	47.5%
Lasso	0.0369	(0.0002)	47.6%
GBT:time-path	0.0363	(0.0002)	48.5%

C 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 however locally flat, which means that finding an optimum is difficult. To address this problem we first generate 30 candidate points using the following global optimization in parallel, using a 100 individuals with one common set of constant 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. This time using a different fixed realization of the random variables. The output of this local optimization is then returned as one candidate solution.

Once these 30 candidate points are found, they are each evaluated using a population size of 10,000, with a new fixed realization of the random variables for all 30 candidates. The best of these parameters are then returned as the solution.

D The relative influence of game parameters and learning

The main learning model incorporates both a direct effect from the game parameters and learning effect in the expression for initial round cooperation

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

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.

We consider the last supergame of each experimental session. As a first step we can look the variation in both $e_i(s)$ and $\alpha + \beta \cdot \Delta^{RD}$. We consider the actual data and a simulated data set with 16 participants in each session. We can compare both the variation between individuals and between the average values of sessions. Since these two values enter the expression in the same way, they are directly comparable.

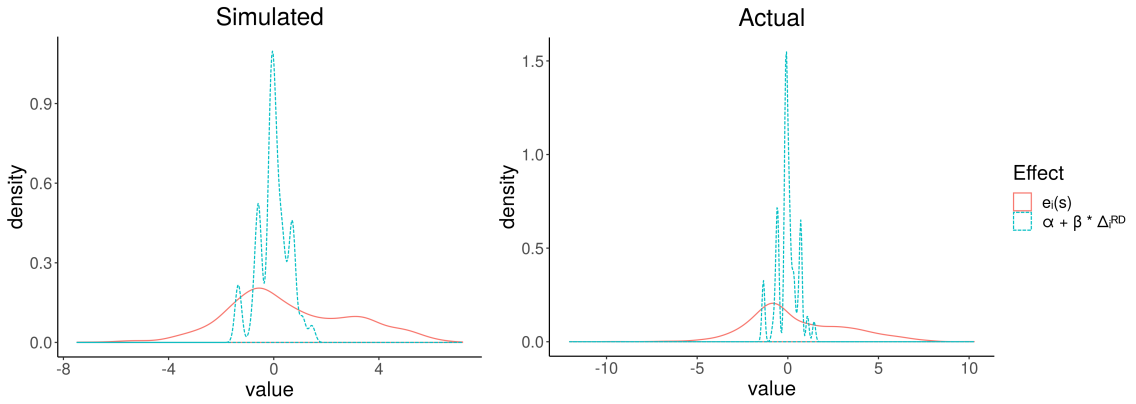

Figure 14: Across individuals variation in experience $e_i(s)$ and direct effect of the learning parameters $\alpha_i + \beta_i \cdot \Delta^{RD}$. Evaluated on the last supergame for each session. Simulated data to the left and actual data to the right.

The variation in the learning effect is much larger than the game parameter effect, and thus has a greater influence on the variation in behavior.

We can also consider the across session variation, averaging the $e_i(s)$ over the individuals in each session.

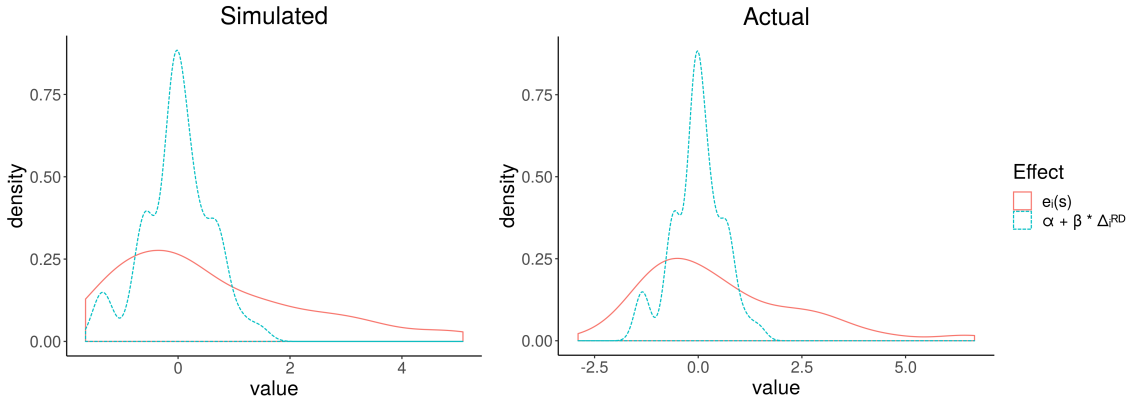

Figure 15: Across sessions variation in experience $e_i(s)$ and direct effect of the learning parameters $\alpha_i + \beta_i \cdot \Delta^{RD}$. Evaluated on the last supergame for each session. Simulated data to the left and actual data to the right.

Since there is considerable variation in $e_i(s)$ between the individuals in a session, but Δ_i^{RD} is the same for everyone in a session, the differences are smaller here than for the individual data. But we still see that the variation in the direct learning effect is larger than the variation in the direct game parameter effect.

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.

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

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

Due to the strong correlation between Δ^{RD} and $e_i(s)$ and the fact that the influence is not linear, the logistic function is more sensitive around 0, we consider

what is essentially the Shapley value of the two effects[21]. 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.

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

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

This can of course be done on a session level instead of an individual level, where the probabilities $p_i(s)$ are first averaged for each session. In the table below we see the results.

Data	$Var(p\|e,\Delta^{RD})$	$Var(p\|e)$	$Var(p\|\Delta^{RD})$	Rel Imp $e_i(s)$	Rel Imp Δ^{RD}
Simulated individual	0.131	0.109	0.019	84.4%	15.6%
Actual individual	0.137	0.119	0.018	86.8%	13.2%
Simulated session	0.079	0.052	0.019	70.8%	29.2%
Actual session	0.078	0.055	0.019	72.7%	27.3%

Table 12: Relative importance measures.

We see that in both the simulated and actual data, $e_i(s)$ is responsible for roughly 85% of the variation in predicted individual behavior, and roughly 70% of the variation in initial round cooperation between sessions.

As a test of whether we should believe these estimates, we see how well they fit with behavior in the last supergames. According to our model, in a logistic regression of initial round cooperation on $e_i(s)$ and $\alpha + \beta\Delta_i^{RD}$, both should have a coefficient of 1, while the intercept should be 0.

[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.

Table 13: Logistic regression on actual individual last supergame initial cooperation. According to our model both these coefficients should be 1 and the intercept 0.

	Dependent variable:
	Actual last supergame initial cooperation
$e_i(s)$	0.806^{***} (0.046)
$\alpha + \beta\Delta_i^{RD}$	1.025^{***} (0.138)
Constant	0.010 (0.069)
Observations	1,734
Note:	*p<0.1; **p<0.05; ***p<0.01

As we see, the coefficient and the intercept are close to the values implied by our model. The coefficient for $e_i(s)$ is slightly lower at 0.8 instead of 1, while the coefficient for $(\alpha + \beta\Delta_i^{RD})$ and the intercept are almost exactly the implied values.

In summary, our model implies that most of the variation in initial round cooperation is driven by experience and not directly by the game parameters. Furthermore, actual behavior in the last session of the supergame is consistent with that result.

E The Pure Strategy Learning Model

Here we outline the belief learning model from Dal Bó and Fréchette (2011), and our across-treatment generalization. Individuals are assumed to choose between TFT or AllD at the beginning of each supergame. The decision is made via a logit best-reply based on the individuals beliefs about how likely a partner is to play TFT or AllD, and the implied expected payoffs.

The beliefs are tracked by the two values B_{is}^C and B_{is}^D, where i is the individual and s is the supergame. Since only two pure strategies are considered, and they prescribe different actions in the initial round of a supergame, the initial-round actions reveal the partner's strategy. The belief values are updated according to

$$B_{is+1}^a = \theta B_{is}^a + \mathbb{1}\{a_{-i}(s) = a\}$$

where $a_{-i}(s)$ denotes the initial round action taken by the partner of individual i in supergame s, and θ captures recency in the beliefs. Given those two belief values, the belief that the partner will play TFT in supergame s is given by $B_{is}^C/(B_{is}^C + B_{is}^D)$.

Let $u^a(\text{TFT}), u^a(\text{AllD})$ denote the expected payoff from taking action a in the initial round if the partner is playing TFT and AllD respectively. Now given the beliefs and those values, the expected value of each choice is given by

$$U^a_{is} = \frac{B^C_{is}}{B^C_{is} + B^D_{is}} u^a(TFT) + \frac{B^D_{is}}{B^C_{is} + B^D_{is}} u^a(AllD) + \lambda_{is}\epsilon^a_{is}$$

where ϵ^a_{is} follows a type I extreme value distribution $\lambda_{is} = \lambda^F_i + (\phi_i)^s \lambda^V_i$. is a sensitivity parameter. This gives the following probability of subject i playing a in the initial round of supergame s, and thereafter following the according pure strategy,

$$p^a_{is} = \frac{\exp\left(\frac{1}{\lambda_{is}} U^a_{is}\right)}{\exp\left(\frac{1}{\lambda_{is}} U^C_{is}\right) + \exp\left(\frac{1}{\lambda_{is}} U^D_{is}\right)}.$$

E.1 Alternative Specifications

Since this model assumes noiseless behavior after the initial round, one possible improvement would be introducing such noise. Therefore, we add a variable ε_i, so that the individual takes the prescribed action with probability $1 - \varepsilon_i$. Otherwise the model remains the same, including using the theoretical values for the value of TFT against TFT etc. While this does improve predictions, it is not so much it changes any conclusions.

In a similar spirit, it could be the case that the pure strategy learning model is only worse in the later rounds, but predicts initial round play well. We therefore also consider the prediction task of only predicting the evolution of initial round play. In the table below we see the out of sample performance for the two different tasks. The difference in predictive performance is no lower in the initial rounds, so that cannot explain the difference in out of sample predictions.

	Time path		Average Cooperation	
Model	Initial	Subsequent	Initial	Subsequent
Pure strategy without ε	0.0364	0.0436	0.271	0.0240
	(0.0004)	(0.0002)	(0.0002)	(0.0002)
Pure strategy with ε	0.0366	0.0396	0.283	0.0217
	(0.0005)	(0.0003)	(0.0002)	(0.0001)
Initial round learning	0.0289	0.0357	0.0200	0.0174
	(0.0002)	(0.0002)	(0.0001)	(0.0001)

Table 14: Comparison of out of sample MSE at initial and non-initial rounds.

F Alternative Hypothesis Tests

As an alternative to the standard errors presented in the text, 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.

Model	Difference	t-test p-value	signed test
OLS on Δ^{RD}	-0.0016	p=0.0012	p=0.0012
OLS	-0.0010	p=0.0483	p=0.0898
Lasso	0.	p=0.8961	p=0.5230
All memory-1 learning	0.0006	p=0.0704	p=0.060

Table 15: Differences and paired significance test with the main learning model for the average cooperation prediction task.

In pairwise tests of the MSE for the predictions of average cooperation, the only consistently significant difference is with the linear function of Δ^{RD}. This tells us that the main learning model is capturing regularity in average cooperation not captured by considering Δ^{RD} alone. However, the difference in MSE between the learning model and the full OLS is only marginally statistically significant.

Model	Difference	t-test p-value	signed test
OLS on Δ^{RD}	-0.0057	p<0.001	p<0.001
OLS	-0.0028	p<0.001	p<0.001
GBT:time-path	-0.0025	p<0.001	p<0.001
All memory-1 learning	0.0024	p<0.001	p<0.001

Table 16: Differences and paired significance test with the main learning model for the time-path prediction task.

On the time path, however, all the differences considered are significant at the 0.0001 level. So initial round learning model is significantly better at predicting the time-path of average cooperation than all model-free prediction methods we consider, but allowing for learning at all memory-1 histories significantly improves it further.

G Evaluation of the procedure on simulated data

To test our estimation approach, we simulate the data using three different models: initial round learning, initial round learning with individual parameters drawn from a normal distribution, and lastly the pure strategy belief-learning model. From each of these models we generate 10 different data sets that mimic the data we have. Each session is simulated with an actual sequence of supergame lengths, and with 16 participants in each session. On each of these 10 different data sets we perform the estimations from the main text, and report averages and standard deviations from these 10 estimations. For the initial-round learning model without additional noise, we take the parameters from table 4, i.e. $\alpha = -0.38, \beta = 3.09, \rho = 0.92, \beta_{reinforce} = 0.13$. The results can be seen in table 17.

Model	Time-path	Average cooperation
OLS on Δ^{RD}	0.0315 (0.0011)	0.0080 (0.0009)
OLS	0.0269 (0.0014)	0.0070 (0.0012)
GBT	0.0239 (0.0012)	0.0055 (0.0009)
Lasso	0.0269 (0.0014)	0.0070 (0.0013)
Pure strategy belief learning	0.0390 (0.0018)	0.0137 (0.0015)
Initial round learning	0.0201 (0.0008)	0.0040 (0.0007)

Table 17: Averages and standard deviations of 10-fold cross validation on 10 different simulated data sets of the initial round learning model without additional noise.

We see that initial-round learning has the lowest average MSE for both prediction tasks. For the time-path task, we can be quite sure to get a difference in any given data set. For the average cooperation, we see that the difference to the pure strategy belief learning model, and the OLS on only Δ^{RD} are likely to happen in any given dataset. However, the difference with the best black-box prediction is quite small relative to the standard deviation.

Next we relax the assumption that that each individual has the same parameters, and let the parameters for each individual be drawn from a normal distribution where $\alpha \sim N(-0.38, 0.5)$, $\beta \sim N(3.09, 1)$, $\rho \sim N(0.92, 0.05)$, and $\beta_{reinforce} \sim N(0.13, 0.05)$. The standard deviations were chosen to give what we thought was reasonable variation. Table 18 describes the results from this exercise.

Model	Time-path	Average cooperation
OLS on Δ^{RD}	0.0297 (0.0008)	0.0075 (0.0008)
OLS	0.0257 (0.0011)	0.0070 (0.0009)
GBT	0.0235 (0.0011)	0.0057 (0.0006)
Lasso	0.0257 (0.0011)	0.0069 (0.0010)
Pure strategy belief learning	0.0383 (0.0017)	0.0138 (0.0021)
Initial round learning	0.0210 (0.0008)	0.0048 (0.0004)

Table 18: Averages and standard deviations of 10-fold cross validation on 10 different simulated data sets of the initial round learning model with noisy individual parameters.

While the MSEs are slightly higher and the differences slightly smaller, there are no substantial differences compared to the simulations without noise.

Lastly, we want to assure that we can distinguish between our learning model and a different learning model. For this purpose we consider the pure strategy belief learning model, with a detailed description in appendix E, with estimated on the real data.

Model	Time-path	Average cooperation
OLS on Δ^{RD}	0.0420 (0.0015)	0.0088 (0.0011)
OLS	0.0304 (0.0015)	0.0068 (0.0009)
GBT	0.0287 (0.0015)	0.0066 (0.0014)
Lasso	0.0304 (0.0015)	0.0071 (0.0010)
Pure strategy belief learning	0.0219 (0.0013)	0.0044 (0.0008)
Initial round learning	0.0379 (0.0019)	0.0103 (0.0012)

Table 19: Averages and standard deviations of 10-fold cross validation on 10 different simulated data sets of the pure strategy belief learning model.

The table shows that we would be able to detect that if the behavior was given by the pure strategy belief learning model instead of the initial round learning model.

Considering the averages and standard deviations from these estimates we can draw three main conclusions. Firstly, on the time-path problem we should expect the true model to outperform all of the alternatives, even if there is individual variation in the parameters. Secondly, for the average cooperation problem, we can expect the true model to outperform the OLS on Δ^{RD}, and the other model. Lastly, for the

average cooperation problem, the difference between the best performing black-box algorithm and the true model is quite small in relation to the standard deviation, which suggests there is clear possibility of overlap.

H Maximum likelihood estimation of next action

In the main text we focus on the prediction errors of the estimated models of the next action taken by individuals, since this allows for straightforward comparisons between models of different complexities. However, since it is more common in the literature to use to consider the likelihoods instead of predictive abilities, we report these likelihoods for completeness. For every history $h_i(t)$ the behavior of type j is captured by a function $\sigma^j : \mathcal{H} \to [0,1]$ that takes a history and assigns a probability to cooperate. Each model comes with set of parameters. We will go through the different models in the following subsections, but first present the general estimation procedure.

If we let $a_i(t) \in \{-1, 1\}$ denote the action taken by individual i at time t, the likelihood of the observed behavior for participant i if she was of type σ^j with parameters is given by

$$\Pr_i(\sigma^j|\theta_j) = \prod_{t \in T(i)} \sigma^j(h_i(t))^{\mathbb{1}\{a_i(t)=1\}}(1 - \sigma^j(h_i(t)))^{\mathbb{1}\{a_i(t)=-1\}}.$$

Let $\theta = (\theta^j)_{j=1}^J$ denote the parameters of the different types, and let $\phi \in \Delta(J)$ denote their relative share. A is then a pair $m = (\theta, \phi)$, and its likelihood is

$$\mathcal{L}(m|\theta, \phi, I) = \sum_{i \in I} \log \left(\sum_{j=1}^{J} \phi^j \Pr_i(\sigma^j|\theta_j) \right).$$

The model is then estimated by maximum likelihood.

Our main learning model only has four parameters per type, and these four parameters are the same across treatments. In comparison, the pure strategy model incorporates 12 different pure strategies, each with a different mistake probability, and these are estimated separately for each of the 32 treatments. If we were to directly compare the pure strategy model's loglikelihoods with the initial round learning model's loglikelihoods, we would be comparing a model with 736 parameters and one with 4.

To make the comparison more meaningful, here we consider the models estimated separately on each treatment as well as on the overall data, and we include BIC values to compensate for model complexity.

We consider three versions of each model (except the 12-type pure strategy model): A single type, a single type plus an AllD type, and three types. Since the first period learning model does not include AllD as a subset, we also consider a version with three first period learning types and one AllD.

In table 20 we see the loglikelihoods, estimated using H, of the different models, estimated and evaluated on the full supergames, and in table 21 evaluated on the last third of the supergames in each session.

Model	N	AllD	Estimated on	Loglikelihood	BIC
Pure	12		Each Treat	-47625	101137
Mixed	1		Each Treat	-51712	104703
	3		Each Treat	-42229	88809
	1		All Treat	-52828	105775
	3		All Treat	-43679	87739
Initial round learning	1		Each Treat	-46326	93164
	1	Yes	Each Treat	-43451	87926
	3		Each Treat	-43542	89131
	3	Yes	Each Treat	-41891	87366
	1		All Treat	-47821	95689
	1	Yes	All Treat	-45124	90320
	3		All Treat	-44575	89317
	3	Yes	All Treat	-43303	86843
Initial round learning	1		Each Treat	-45462	92716
with flexible	3		Each Treat	-40039	85966
mixed strategy	1		All Treat	-47560	95168
	3		All Treat	-41878	83922
Full learning	1		Each Treat	-44368	90527
	3		Each Treat	-39543	84974
	1		All Treat	-47149	94345
	3		All Treat	-41529	83225

Table 20: Maximum likelihood log-likelihoods evaluated on the complete set of supergames.

Model	N	AllD	Estimated on	Loglikelihood	BIC
Pure	12		Each Treat	-11314	27638
Mixed	1		Each Treat	-14186	29462
	3		Each Treat	-10535	24774
	1		All Treat	-14680	29467
	3		All Treat	-11241	22826
Initial round learning	1		Each Treat	-12125	24686
	1	Yes	Each Treat	-11273	23418
	3		Each Treat	-11168	24079
	3	Yes	Each Treat	-10663	24376
	1		All Treat	-12731	25505
	1	Yes	All Treat	-11898	23862
	3		All Treat	-11440	23031
	3	Yes	All Treat	-11167	22550
Initial round learning	1		Each Treat	-11717	24960
with flexible	3		Each Treat	-10095	25201
mixed strategy	1		All Treat	-12597	25236
	3		All Treat	-10789	21728
Full learning	1		Each Treat	-11210	23945
	3		Each Treat	-9581	24174
	1		All Treat	-12506	25056
	3		All Treat	-10323	20796

Table 21: Maximum likelihood log-likelihoods evaluated on the last third of the supergames.

As shown in the tables above, these maximum likelihood results are consistent with the primary analysis. According to the BIC, the best model is the learning model that extends to all memory-1 histories, while the pure strategies model is one of the worst. However, we still achieve relatively good performance with the simpler learning model that keeps behavior after the initial round constant across treatments and individuals, especially if we include AllD.

We see the same ordering as for the predictions: mixed strategies are better than pure strategies, first period learning plus AllD is better than mixed strategies, fist period learning with flexible mixed strategies better still, and the best is the full learning model.

We also see that we accurately capture the between treatment variation within our models. The loglikelihood is often similar for the models estimated jointly for all treatment, with logistic functions of Δ^{RD} capturing the variation between treatments, and the ones estimated separately for each treatment. And the lowest BIC is given by such joint estimation.