Predicting Cooperation with Learning Models

Drew Fudenberg [1] Gustav Karreskog [2]

[1]MIT

[2]Uppsala University

Politecnico di Torino, November 10 2022

Equilibrium theory of repeated games

- Static outcomes with low discount factors.
- Folk theorems when players are patient.

- Repetition does not rule out the static equilibria, can only add equilibria and not take them away.

- *Intuition:* people tend to cooperate when this is an equilibrium.

- Economic analyses often assume that people play the equilibrium with the most cooperation, but this is a poor fit for observed behavior in the laboratory.

Laboratory Evidence: Repeated PD with Perfectly Observed Actions

- In Roth and Murnighan [1978] and Murnighan and Roth [1983], subjects only played one iteration of the indefinietly reepeated prisoner's dilemma (PD).
- Effect of discount factor *mostly* but not that much cooperation even when discount factor high.
- *Reason*: Subjects only played the repeated game once, didn't get a chance to learn from experience.
- In Dal Bo [2005], subjects played 7–10 iterations of the repeated game, with a different partner each time.
- Much more effect of the discount factor than when theu just play once.
- Launched a revival of repeated game experiments, mostly on variations of the PD.

- Past work has focused on trying to learn how individuals play the game. We consider different problem: How overall cooperation rates depend on parameters.
- We consider the problem of predicting cooperation rates out of sample, using data from past experiments on the repeated prisoner's dilemma with observed actions.
- No theoretical results; we use simulations of a simple learning model to make our predictions.
- Use machine learing performance as a benchmark.
- The simplest version of the model, where learning only affects choices in the initial round of each supergame, performs at least as well on our data as more complicated models and machine learning algorithms.
- Our results help explain past findings on the impact of risk dominance considerations, and also suggest some modifications to them.

Preliminaries

- Data come from the 12 papers covered in the Dal Bó and Fréchette [2018] meta-analysis and 5 papers published since then: 161 experimental sessions and 2,612 participants.
- Participants played a sequence of repeated prisoner's dilemma games with perfect monitoring. The game parameters were held fixed within each session.
- Randomly chosen partners and a random stopping time, so the discount factor δ determines the probability $(1-\delta)$ 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."
- Normalize the payoff to joint cooperation to 1 and the payoff to joint defection to 0; this is w.l.o.g. for agents who maximize expected utility.

$$\begin{array}{ccc} & C & D \\ C & R, R & S, T \\ D & T, S & P, P \end{array} \qquad \begin{aligned} T > R &> P > S, \\ 2R &> T + S \end{aligned}$$

Normalize: substract P from all the payoffs, divide by $R - P$

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

$g - l < 1$

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

- "Cooperate every round" is the outcome of a subgame-perfect equilibrium if and only if

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

- Note that the loss l incurred to (C, D) does not enter in to this equation!
- Little experimental support for the idea that players cooperate if $\delta > \delta^{\mathsf{SPE}}$.
- Cooperation in repeated game experiments can be better predicted by measures that reflect uncertainty about the opponents' play.

- A strategy is risk dominant in a 2x2 game if it is the best response to a 50-50 randomization.
- 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 RD-difference is

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

- No good reason to think people only consider Grim and Always Defect, but Δ^{RD} proves to be a useful composite parameter or "feature."
- Unlike $\delta^{\text{SPE}} = g/(1+g)$, Δ^{RD} depends on l as well as g, which makes sense if people aren't sure how their partners will play.

Overview: Prediction Tasks

- We consider two prediction tasks: Predicting average cooperation in a given session, and predicting the time path of cooperation over the course of a session.

- Predicting time paths lets us make predictions about what cooperation levels would be if sessions were longer.

- And predicting the time paths is also a good way to predict the average: when predicting the average cooperation level in a session, each session is a single observation, while when predicting the time path of cooperation, a data point is the average (across participants) cooperation level on each round of each supergame.

The IRL-SG Model

- The simplest model we consider is IRL-SG: "Initial Round Learning-Semi Grim."
- In this model, learning only influences how agents play in the initial round of each supergame.
- Subsequent play in the supergame is determined by a "semi-grim" strategy.
- With a semi-grim strategy, the action played only depends on the outcome of the previous round, so it is "memory-1."
- The restriction to memory-1 strategies is motivated by past work, and also by our machine learning analysis
- And play after CD is assumed to be the same as after DC.
- The IRL-SG assumes the same semi-grim strategy is used by all individuals in all treatments.
- So we need to estimate 3 parameters to pin it down.

Initial Play in the IRL-SG

- We allow initial-round cooperation in supergame s to depend on Δ^{RD}.
- And we also allow it to depend on each individual i's past experiences $e_i(s)$— this is where learning comes in.
- Specifically, we assume that initial-round cooperation $p_i^{initial}(s)$ is given by :

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

- Cooperation or defection in the initial round is reinforced via the updating of e_i:

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

where $a_i(s)$ is -1 if i played D in the initial round of supergame s, and 1 if i played C, and $V_i(s)$ is the *total* payoff i received in supergame s.

- λ determines the strength of learning.

- To initialize the system we set $e_i(1) = 0$, so in the initial round of the first supergame all individuals in a session randomize in the same way.
- Given a simulated population, we can calculate either average cooperation or the time path of cooperation.
- Key:Our predictions don't use endogenous data like actions played and payoffs received, they only depend on the game parameters and realized game lengths.
- We estimate the learning model based on the time path of cooperation, even when predicting average cooperation. That is, we find the parameters that best predict the *time path of cooperation* in the training set, and use those parameters to predict both the average cooperation and the time path of cooperation in the test sets.
- This lets use more of the data to estimate the model.

Estimation

- We estimate model parameters numerically on 10 folds, holding fixed the realizations of the random variables as we tune the parameters.
- We evaluate models by their 10-fold cross-validated mean squared error (MSE).
- Our method doesn't have performance guarantees, but it performs well on on simulated data; in particular it can distinguish our IRL-SG model from pure-strategy reinforcement learning.
- Train/test splits are 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 MSE, we do 10 different such 10-fold cross-validations. This results in 100 different MSE values from which we estimate the standard errors.

Machine Learning Algorithms

- We used ML algorithms (and OLS) to make predictions.

- The ML algorithms include Lasso, Support Vector Regressions (SVR), and Gradient Boosting Trees (GBT). (Tried others that didn't work well.)

Features

- For average cooperation we used Δ^{RD}, δ, g, l, total #rounds, #supergames, an indicator for $\Delta^{RD} > 0$, summary statistics for the difference between expected and realized supergame lengths, and some interactions.

- For predicting time paths, added current supergame, current round, and an indicator for the initial round. Details

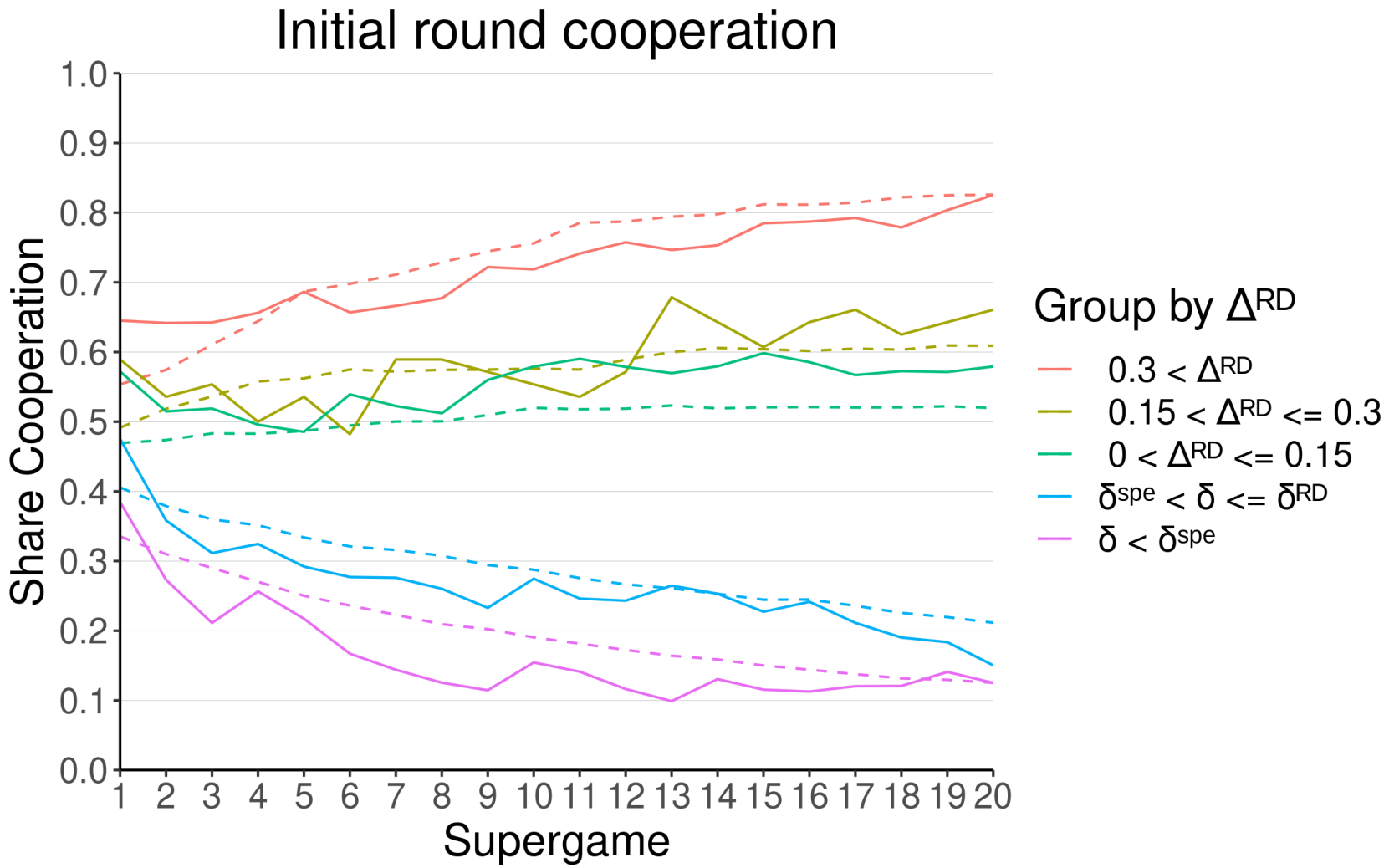
- With enough data, ML algorithms can figure out interactions or composite variables by themselves. With limited data, including these as features can help, so we can't rule out that other features could yield better predictions.

Predicting average cooperation

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

- Our learning model (slightly) outperforms the ML algorithms because it better predicts the influence of realized supergame lengths—in particular, the model predicts that there is more cooperation when the realized supergames are long.
- If we remove supergame lengths from the learning model and ML algorithms, they both have prediction error $.0158$.
- Similar results for the time-path problem. Time-path table

Actual and predicted initial round cooperation



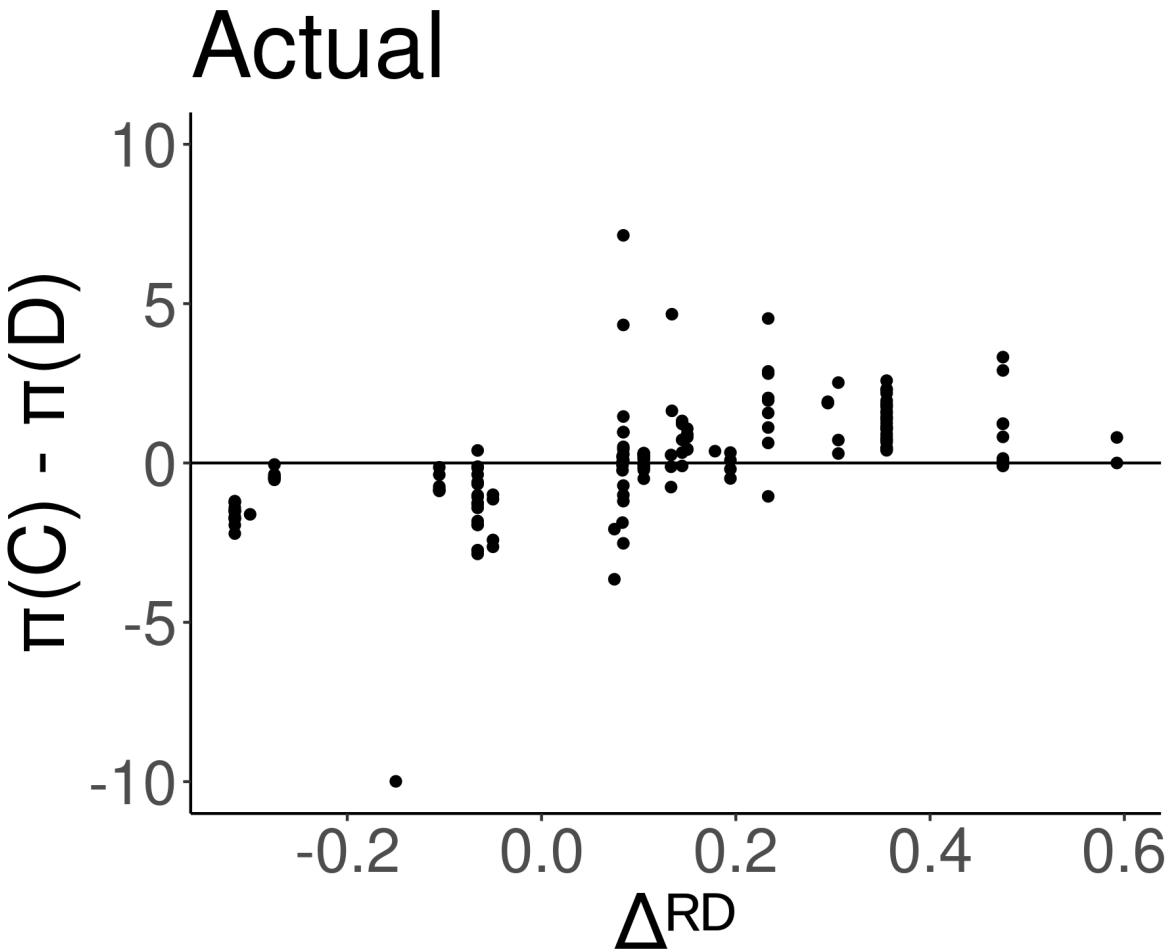
Actual (solid line) and out of sample predicted (dashed line) initial-round cooperation by supergame for sessions of at least 20 supergames.

Interpreting λ

- The estimated learning rate ($\lambda = 0.182$) implies a strong learning effect.

- With the estimated parameters, approx 88% of the between-treatment variance in predicted cooperation in the initial round of the last supergame in a session is driven by learning.

Δ^{RD} and Learning



Average **empirical** difference between total payoff in supergames where the participant cooperated and defected in the first round. Each dot corresponds to one experimental session.

Δ^{RD} and Learning

- For $\Delta^{RD} < 0$, initial-round defection is reinforced in all but 1 session.
- For positive but low values of Δ^{RD}, the difference in reinforcement $\pi(C) - \pi(D)$ is centered around 0, so cooperating and defecting are on average reinforced equally.
- This helps explain why there aren't clear time trends in the sessions where $0 < \Delta^{RD} < 0.15$.
- Simulations of our model show a similar pattern.

Adding one parameter

- Our baseline IRL-SG model has 6 parameters: α and β determine how Δ^{RD} influences initial play, λ determines the strength of reinforcement, and σ_{CC}, σ_{DD},and $\sigma_{CD=DC}$ are the memory-1 mixed strategies.

- The following keep those parameters and add one more:

  - *Recency effect*: $e_i(s) = \lambda \cdot a_i(s-1) \cdot V_i(s-1) + \rho \cdot e_i(s-1)$, where $\rho \in [0, 1]$ discounts previous experiences.

  - *Flexible reinforcement threshold*: cooperation is reinforced for payoffs greater than τ instead of greater than 0.

  - *Learning with memory-1*: drop the requirement that $\sigma_{DC} = \sigma_{CD}$; 7 parameters

Variations with more parameters

- *Initial round learning with flexible memory-1*. Allow play after each memory-1 history to depend on Δ^{RD}; 11 parameters.

- *Learning at all memory-1 histories*: Cooperation probabilities are updated at the beginning of each supergame, and remain constant in its subsequent rounds; 11 parameters.

- *Learning at all memory-1 histories with two rates*: Separate learning rates λ for initial-round actions and non-initial actions; 12 parameters.

- *Exogenous heterogeneity*: Two types of IRL-SG agent, and a parameter that determines their population shares. This is the only variant that has better outsample performance, but the improvement is slight, and it has 13.

- *IRL-SG and AllD*: Two types, one follows IRL-SG and the other Defects with constant probability $1 - \varepsilon$. (*This variant is suggested by the obervation that some subjects seem to defect almost all the time.)*

Alternative Models from the Literature

- Dal Bó and Fréchette [2011] estimate a learning model where all participants make a noisy choice between either Tit for Tat (TFT) or AllD, with expected payofffs computed as in fictitious play with recency.

- They estimate insample fit. To make cross-treatment predictions we let initial beliefs depend on Δ^{RD}, adding 2 parameters. We also allow for "trembles."

- And we consider a reinforcement learning model with TFT, AllD and Grim.

- Most of the more flexible learning models performed a bit worse than the IRL-SG; the two-type IRL-SG did slightly better.

Extrapolating to Longer Experiments

- We are interested in what would happen over a longer time scale than feasible in the lab.
- We use our learning model to make predictions about that.
- But first we test how well we can extrapolate from the first half of the sessions to the second half: If we can't do that well it's hard to be very confident about extrapolations to out-of-sample session lengths.
- As before, each session is in either a training or a test fold but not both.
- But now we use the first halves of the training sessions to predict play in the second halves of the test sets.

Prediction loss (MSE) from estimating on first half of a session and evaluating on its second half.

Model	2nd half MSE for avg C
Constant	0.0695 (0.0055)
SVR	0.0285 (0.0030)
Lasso	0.0284 (0.0030)
OLS on Δ^{RD}	0.0281 (0.0032)
GBT	0.0266 (0.0027)
IRL-SG	0.0220 (0.0026)

- The learning model is better at extrapolating to the second halves of the sessions than our black-box algorithms or simple OLS, and the difference is statistically significant.
- Could be due to our particular ML implementations, but a more structured model that encodes some intuition or knowledge about the problem domain can sometimes better extrapolate to related prediction problems.

- The performance of the IRL-SG in extrapolating from 1st halves of sessions to their second halves gives us some confidence in our extrapolation to cooperation rates in longer sessions than we see in our data.

- When we do this we see very wide 90% intervals for intermediate values of Δ^{RD}, less variation when $\Delta^{RD} < 0$ or $\Delta^{RD} > 0.3$

- We now extrapolate long-run play in the six treatments in Dal Bó and Fréchette [2011].

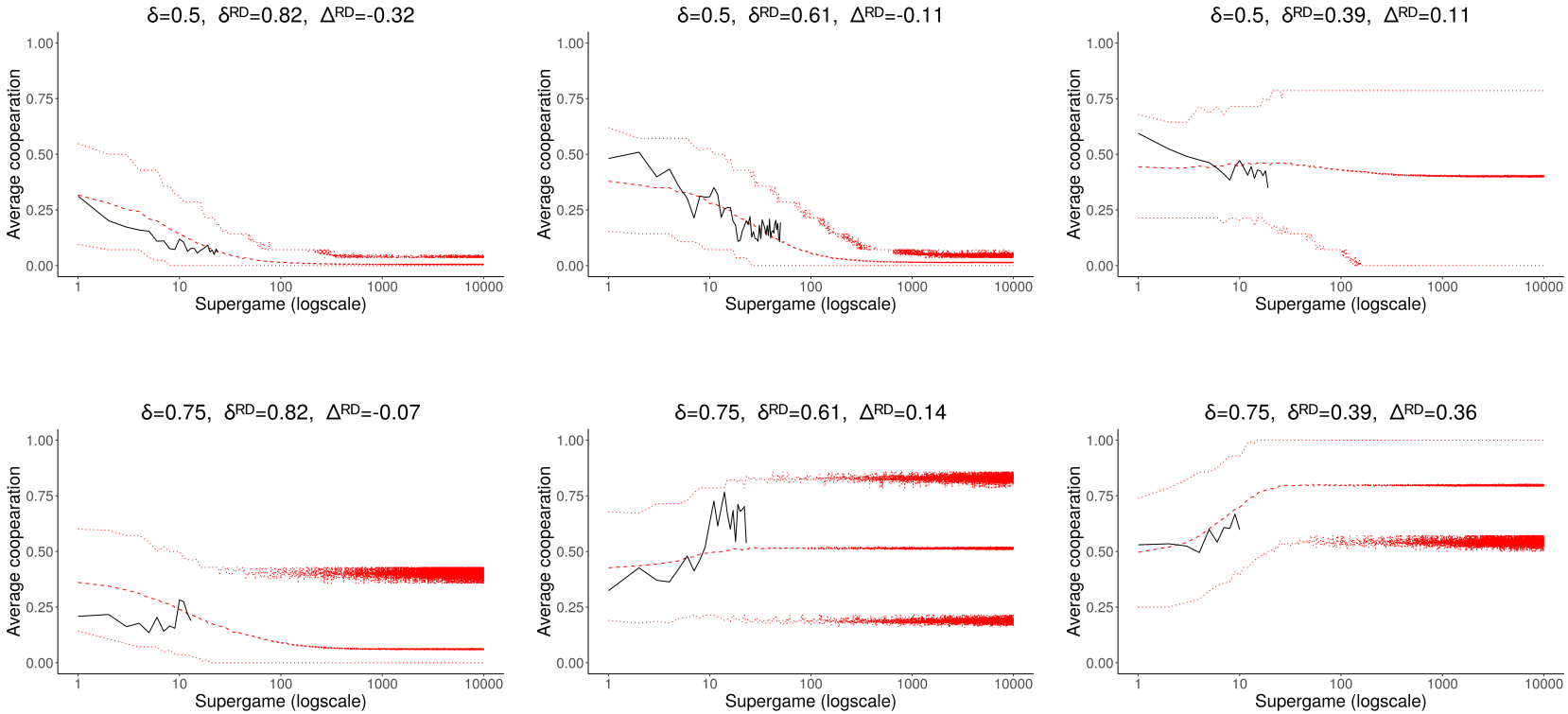
- For each treatment, 1,000 populations with 14 participants were simulated for 10,000 supergames, with randomly drawn supergame lengths.

- Randomness in both behavior and the differing experiences can lead to substantially different outcomes.

  - For $\Delta^{RD} < 0$, we see less than 50% cooperation.

  - For $\Delta^{RD} = 0.11$, even after 10,000 supergames the 90% interval goes from 0% cooperation to 79%.

  - For $\Delta^{RD} = 0.36$ relatively certain prediction of high rates of cooperation.



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 1,000 simulated populations, the dotted red lines depict the middle 90% interval.

- Wide 90% intervals for intermediate values of Δ^{RD} due to the randomness of behavior and small population size.

- Randomness comes from random initial play in a finite population: When $\Delta^{RD} = 0.11$ and population size is 100, the 90% interval is [15, 64]; with 1,000 participants it is [24, 56].

- Randomness also comes from the realized supergame lengths: If population size is 1,000 and all of the simulated supergames have their expected number of rounds, the 90% interval shrinks to [44, 49].

- The intervals are smaller in treatments where Δ^{RD} is more extreme in either direction.

Discussion

- The key to predicting cooperation in a given match is the prediction of play in the initial round.
- Initial-round play depends on the game parameters and on experience in previous matches.
- Can predict fairly well with a simple learning model that holds play fixed except in the initial round of each supergame, and has only 6 parameters
- The learning model with a single type has endogenous heterogeneity; little gain by adding exogenous type-heterogeneity—neglecting learning might lead researchers to overemphasize heterogeneity in participants. (Paper reports a similar finding for predicting the next action played.)

- Why does there seem to be little adjustment to play at non-initial rounds?

- *Conjecture*: there is a cognitive cost associated with learning from experience and adjustment to game parameters, so we should expect learning and adjustment to happen where the relative payoff is the greatest, and this is in the initial round.

- To test this, we assume all other individuals behave according to our estimated IRL-SG model, and calculate the potential gain from learning and adjustment at different histories.

- The best IRL-SG does much better than best model w/o learning, and learning at all histories yields only small improvement, which supports the conjecture.

- The main way Δ^{RD} influences cooperation is through the probability of cooperation in the initial rounds of each match.
- Initial cooperation trends up or down depending on whether it is positively or negatively reinforced, which depends on Δ^{RD}.
- Our model also predicts that the values of g and l have an effect that isn't captured by the composite parameter Δ^{RD}: increasing $g - l$ for fixed $g + l$ dampens the effect of any given Δ^{RD}. (There aren't enough experiments with the same Δ^{RD} and different g, l to directly test this prediction.) Details
- Our model lets us predict what average cooperation rates would be with longer lab sessions (assuming the participants did not loose focus on the task).
- Many real-world settings have implementation errors or imperfect monitoring. Not yet enough experimental studies of these games to test cross-treatment predictions. Once there are it would be useful to extend our analysis to them.

One Step Ahead Prediction (OSAP)

- We also considered the problem of predicting an individual's next action $a(t+1)$ conditional on the history $h(t)$.
- Resembles a common approach in the literature: choose parameters of a structural model to maximize the in-sample likelihood of all individual decisions.
- If participants use memory-2 or memory-3 strategies, access to the 3 preceding rounds, instead of just the previous, should improve predictions.
- Most of the time, participants just repeat the previous action.

Models for OSAP

- Allow heterogeneity: We assume there are a finite number strategies or learning rules used in the population, and estimate the parameters and the shares of these strategies by maximum likelihood.

- Each round we calculate the posterior probability of an individual being of each type and make the corresponding predictions.

- Compare performance to a naive benchmark that predicts the previous action taken by the individual, and to the predictions made by a gradient boosting tree.

Model	N types	Loss	Accuracy
Naive		0.383	87.3%
Memory-1 mixed strategy	1	0.343	81.4%
	3	0.245	89.7%
Flexible memory-1	1	0.307	86.3%
	3	0.241	90.2%
IRL-SG	1	0.266	90.4%
	3	0.236	90.8%
Learning at all memory-1	1	0.267	90.6%
	3	0.226	91.3%
GBT with memory-1		0.167	93.7%
GBT with memory-3		0.164	93.8%

Table: Out of sample prediction errors for predicting the next action taken by an individual evaluated on the last third of the supergames. Loss is average negative log-likelihood

OSAP or Time-path?

- OSAP estimated models give noisy and unstable predictions of average cooperation.
- Most individual actions are easy to predict. The "naive" model has an accuracy of 87.3 %, best model has 93.8 %.
- Initial round actions have most of the impact on the time path.
- Time paths of aggregate behavior are what matters for welfare.

Ongoing Research – Learning in Static Games

- Learning in static games with random matching is relatively well understood theoretically; we' d like to better understand the learning rules actually used.

- Current literature focuses on maximizing likelihood of individual decisions, as in the OSAP problem.

- This appoach has low power and biased wstimates (Salmon (2001), Cabrales and Garcia-Fontes (2000), and Wilcox (2006).)

- We will extend the idea of predicting aggregate time paths to study populations who play a static game with anonymous random matching.

- We plan to collect a large new data set, because much of the existing data isn't publicly available and considers a limited set of games.

Thank you!

Additional Previous Work

- Dal Bó and Fréchette [2011] uses the related measure

$$\frac{(1-\delta)l}{1-(1-\delta)(1+g-l)};$$

  it is very correlated with Δ^{RD} and again reflects the role of l.
- [Dal Bó and Fréchette, 2018, 2011; Engle-Warnick and Slonim, 2006] find there is more cooperation in the first round increased if the realized length of the previous supergame is longer than expected.
- Past work also suggests that most participants use memory-1 strategies at least when the PD has perfect monitoring; see e.g. Dal Bó and Fréchette [2018, 2011] and Fudenberg, Rand, and Dreber [2012].
- Romero and Rosokha [2018] and Dal Bó and Fréchette [2019] elicit pure strategies from participants and confirm the finding that a small set of memory-1 strategies are enough to capture most of the strategies used.

Back

Details on g-l

- We have 2 treatments for $\Delta^{RD} = -.05$, with 3 sessions of one parameter constellation and 1 of the other, and 4 treatments for $\Delta^{RD} = .0833$, but only 5 sessions in total.

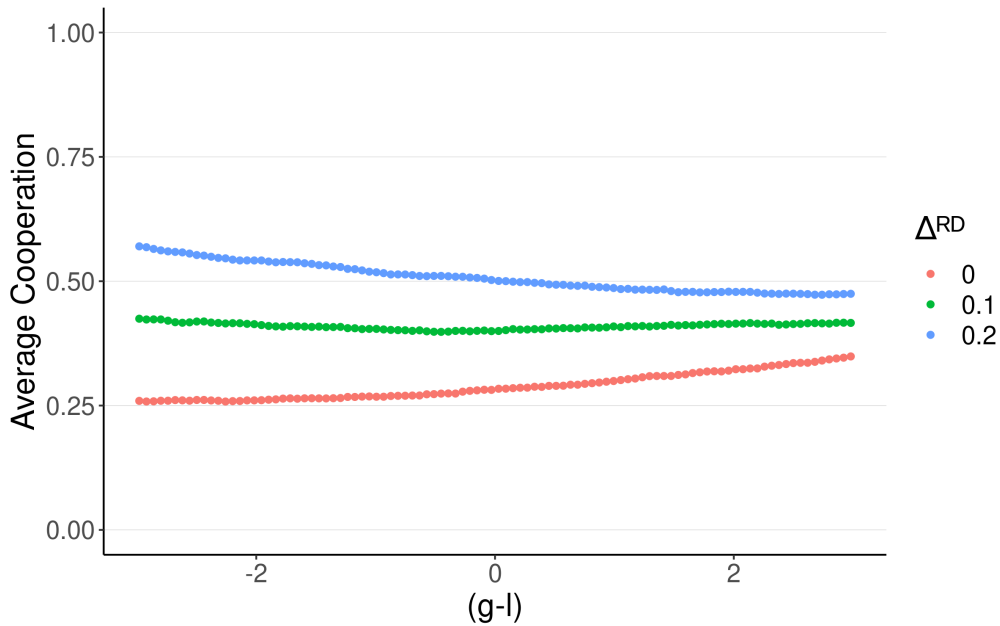


Figure: Predicted average cooperation over 27 supergames for fixed $g + l$ but varying $g - l$.

Back

Details of Features 1

- The expected length of a supergame is $\frac{1}{1-\delta}$. Let $l(s)$ be the realized length (number of rounds) of supergame s.

- The absolute difference between realized and expected is then $l(s) - \frac{1}{1-\delta}$, in relative terms it is $(1-\delta)l(s) - 1$.

- The absolute and relative difference between expected and realized lengths can be averaged over different time-spans, for example over the whole session or the cumulative until supergame s.

Back

Details of Features 2

- **Features for average predictions:** δ, g, l, Δ^{RD}, $rd = (\Delta^{RD} > 0)$, #supergames, #rounds, $rd \cdot \Delta^{RD}$, $rd \cdot (\#\text{supergames})$, $rd \cdot (\#\text{rounds})$, difference in length in the first third of supergames, total difference in length.

- **Features for time-path predictions:** δ, g, l, Δ^{RD}, $rd = (\Delta^{RD} > 0)$, current supergame, current round of supergame, $(round = 1)$, $rd \cdot \Delta^{RD}$, $rd \cdot (\text{current supergames})$, $rd \cdot (\text{current round})$, previous supergame difference in length, cumulative difference in supergame length.

- *We include both the absolute and relative difference for each of the difference measures.*

Back

Time-path Predictions

Model	Time-path
Constant prediction	0.0775 (0.0050)
OLS on Δ^{RD}	0.0398 (0.0025)
GBT:time-path	0.0321 (0.0020)
IRL-SG	0.0309 (0.0020)

Out of sample prediction loss (MSE) of predictions for time path for different learning models.

Back