The Evolution of Cooperation:
The Role of Costly Strategy Adjustments

July 2017

Julian Romero[†]   Yaroslav Rosokha[‡]

Abstract

We study the evolution of cooperation in the indefinitely repeated prisoner's dilemma when it is costly for players to adjust their strategy. Our experimental interface allows subjects to design a comprehensive strategy that then selects actions for them in every period. We conduct lab experiments in which subjects can adjust their strategies during a repeated game but may incur a cost for doing so. We find three main results. First, subjects learn to cooperate more when adjustments are costless than when they are costly. Second, subjects make more adjustments to their strategies when adjustments are costless, but they still make adjustments even when they are costly. Finally, we find that cooperative strategies emerge over time when adjustments are costless but not when adjustments are costly. These results highlight that within-game experimentation and learning are critical to the rise of cooperative behavior. We provide simulations based on an evolutionary algorithm to support these results.

Keywords: Indefinitely Repeated Games, Prisoner's Dilemma, Experiments, Cooperation, Strategies

† Eller College of Management, University of Arizona ● Email:jnromero@gmail.com

‡ Krannert School of Management, Purdue University ● Email:yrosokha@purdue.edu

1 Introduction

A strategy is defined as a complete plan that specifies what a player would do at any point in a given interaction. Since it is a complete plan, it is assumed that a player chooses this strategy before starting the interaction and sticks with it throughout. Implicit in this assumption is that a player has considered every possible contingency and determined an action to be played in each case. This may be problematic in repeated games in which the number of possible contingencies (or histories) grows exponentially as the number of periods of the interaction increases. In this paper, we distinguish between a *permanent strategy* and a *provisional strategy*. In a permanent strategy, the player carefully considers an action after each history. However, specifying such a strategy may be costly (since there are a lot of histories to consider), impossible (if the player is boundedly rational), or not worthwhile (if some histories seem reasonably unlikely). Alternatively, in a provisional strategy, the player sets some contingencies but may not fully consider his action after every possible history. Therefore, if something unanticipated happens, then the player may want to change his strategy during an interaction.[1]

Unanticipated changes to a strategy can carry a substantial cost (either explicit or implicit). For example, in an organization, changes to a strategy may require a board meeting, a shareholder vote, or consultants' advice. Similarly, conducting a vote on a state, county or municipal proposition, which specifies or revises a strategy, carries a substantial cost. Lastly, at the individual level, making changes to the strategy requires additional effort to re-analyze the problem rather than simply continuing with the current strategy. Such decision-making costs have long been accepted as significant (Simon, 1957; Shugan, 1980). In all of these examples, though it is costless to follow the current strategy, revisions during an ongoing interaction are costly. In this paper, we focus on the role of costly strategy adjustments during ongoing interactions and, therefore, pay particular attention to provisional strategies (and refer to these as strategies for simplicity).

An ongoing interaction is typically modeled as a repeated game (also referred to as a supergame), and much work has been done to understand strategies that arise across multiple repeated games (Dal Bó and Fréchette, 2011; Fudenberg, Rand, and Dreber, 2012; Bigoni, Casari, Skrzypacz, and Spagnolo, 2015; Dal Bó and Fréchette, 2015). However, to the best of our knowledge, little work has been done to understand factors that are important within supergames. This paper aims to identify one such factor. Specifically, we examine whether costly strategy adjustments during an indefinitely repeated prisoner's dilemma impact the evolution of cooperation and strategies.

Our experiment contains two treatments – one in which strategy adjustment is costless and one in which it is costly. The two setups are otherwise identical and include a costless strategy construction stage before each supergame begins. Furthermore, in both treatments, subjects can specify equally complex strategies, and we provide a substantial amount of time for them to do

---

[1]Consider the following example of a revision to established strategy as a response to an unanticipated event: "Oil ministers from six Persian Gulf countries are to meet today to set a strategy for dealing with price reductions by North Sea producers and Nigeria. Algeria has also called for an emergency meeting of the 13-member Organization of Petroleum Exporting Countries to try to avoid a price war...." – *New York Times*, Feb 22, 1983.

so. Thus, because the strategy is set ahead of time, the addition of costs during the supergame should not affect the outcome unless players adjust their strategy during the supergame. However, we find different levels of cooperation and different types of strategies when strategy adjustments are costly compared to when they are costless.

We have three main results from our experiments. First, we find higher levels of cooperation when strategy adjustments are costless compared to when they are costly. Second, we find that subjects make more within-supergame strategy adjustments when they are costless but still make changes to their strategies even when they are costly. Finally, in both treatments, players start with strategies that favor defection but learn to play more cooperative strategies only when it is costless. The key takeaway from the experiment is that imposing costs on within-supergame strategy adjustments restricts subjects' ability to experiment and learn, which ultimately hinders the evolution of cooperation.

Our work contributes to three strands of literature. First, we contribute to the growing experimental research that investigates conditions under which cooperation arises in repeated prisoner's dilemma games (Dal Bó, 2005; Duffy and Ochs, 2009; Camera and Casari, 2009; Stahl, 2011; Dal Bó and Fréchette, 2011; Bruttel and Kamecke, 2012; Friedman and Oprea, 2012; Fudenberg, Rand, and Dreber, 2012; Stahl, 2013; Peysakhovich and Rand, 2013; Honhon and Hyndman, 2015; Bigoni, Casari, Skrzypacz, and Spagnolo, 2015; Breitmoser, 2015; Dal Bó and Fréchette, 2015; Embrey, Fréchette, and Yuksel, 2015; Cason, Lau, and Mui, 2016; Romero and Rosokha, 2017). Dal Bó and Fréchette (2016) provide a review of the literature on repeated prisoner's dilemma games. A closely related paper that does not examine the prisoner's dilemma, Embrey, Mengel, and Peeters (2016b), investigates whether strategy revision opportunities affect collusion and strategy choices in indefinitely repeated games of strategic complements and substitutes. They find that allowing a unilateral revision opportunity (at a small cost) affects collusive behavior in games of strategic complements but not in games of strategic substitutes.[2]

Our goal is unique in that we aim to understand the evolution of cooperation and strategies across supergames in an indefinitely repeated prisoner's dilemma in which strategy adjustment is costly within an interaction. Our experiment uses a modified version of the experimental interface of Romero and Rosokha (2017). The interface is designed to run lab experiments with high continuation probability and to elicit strategies from the subjects. The flexibility of the elicitation method allows subjects to create arbitrarily complex strategies, which ensures that they can implement their desired strategy from the beginning of the supergame and not incur any adjustment costs. If subjects do choose to change their strategy, then the interface allows them to make these changes in an intuitive manner in real time during the supergame. We find that when strategy adjustments are costly, the cooperation rate is significantly lower than when it is not.

The second strand of literature to which our paper contributes is the research in game the-

---

[2]Other work has been done to try to understand cooperation in repeated games other than the prisoner's dilemma. These include public goods games (Lugovskyy, Puzzello, Sorensen, Walker, and Williams, 2017); games of strategic complements and substitutes (Embrey, Mengel, and Peeters, 2016a); and trust games (Engle-Warnick and Slonim, 2004, 2006; Casari and Cason, 2009).

ory and experimental economics that investigates learning in repeated games (see Ho, 2008, for a review). Models in which players learn over actions often have a difficult time explaining experimentally observed levels of cooperation. However, models in which players learn over strategies have had more success (Hanaki, Sethi, Erev, and Peterhansl, 2005; Ioannou and Romero, 2014). When modeling strategy learning, it is not clear whether subjects play a fixed strategy for the entire supergame and then adjust their strategy between supergames, or whether subjects adjust their strategies within the supergame. By imposing the strategy adjustment costs within supergames and allowing free modification of strategies between supergames, our experimental interface allows us to investigate whether subjects learn by adjusting strategies within a supergame or only across supergames. We find that this within-game experimentation is critical to cooperation in an indefinitely repeated setting: cooperation is higher when strategy adjustment is costless and lower when plan adjustment is costly.

Finally, we contribute to a third strand of literature that bridges agent-based computational economics and human subject experiments (see Duffy, 2006, for a review). Specifically, we conduct computational simulations based on an established evolutionary algorithm (Holland, 1975) in order to shed light on the difference in learning dynamics that may arise between the two treatments. We implement a simple model in which an agent is viewed as a collection of strategies, and one strategy is chosen to be played. Similar to our experimental design, a cost is incurred when the previously played strategy is adjusted in favor of another strategy. We find that by imposing an adjustment cost, we substantially hinder the likelihood that cooperative outcomes are achieved for a large set of parameter combinations.

The rest of the paper is organized as follows: in Section 2, we present details of the experimental design. In Section 3, we present the data and main results of the lab experiment. In Section 4, we consider several computational benchmarks that provide intuition for the main results of the paper. Finally, in Section 5, we conclude.

2   Experimental Design and Administration

We recruited 138 students for the experiment using ORSEE software (Greiner, 2015) on the campus of Purdue University. Six sessions of the experiment were administered in April 2016, with the number of participants varying between 20 and 26. The experiment lasted for 90 minutes, of which approximately 35 minutes were devoted to the instructions and the quiz. There were no participant identifiers within the game interface, and participants remained anonymous throughout the experiment.

2.1 Experimental Interface

We conducted the experiment using a novel interface that was a modified version of the one used in Romero and Rosokha (2017).[3] The interface allows implementation of the strategy method for long repeated games and has been shown to match data when subjects make choices directly (Romero and Rosokha, 2017). The strategy method has been commonly used in experimental economics, and much work has been undertaken to understand its similarities to and differences from the game method, also known as the direct-response method (Brandts and Charness, 2011). Unlike the literature reviewed by Brandts and Charness (2011), we do not aim to compare the strategy method to the game method. Rather, we focus on environments in which actions have to be implemented through the strategy.

The interface implements the strategy method via a collection of *if-then* rules that the participants construct, and this collection of rules, which we term a *rule set*, makes choices for them. The key feature of the interface is that it allows subjects with minimal programming experience to design complex strategies in an intuitive manner. Furthermore, subjects can condition on as many periods as they want, with virtually no restrictions on the type of strategies that they can construct. Figure 1 presents a screenshot of the interface.
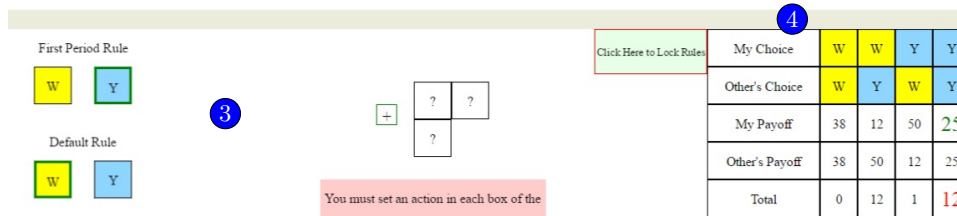
---

[3]More information about the interface, including the code, the instructions and the quiz, are available at `http://jnromero.com/research/strategyChoiceCosts/`.
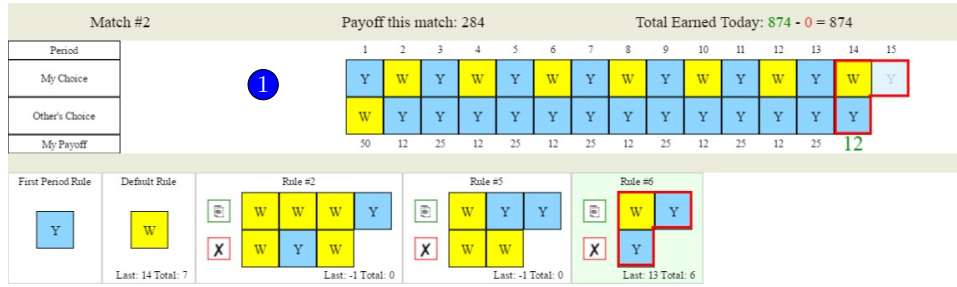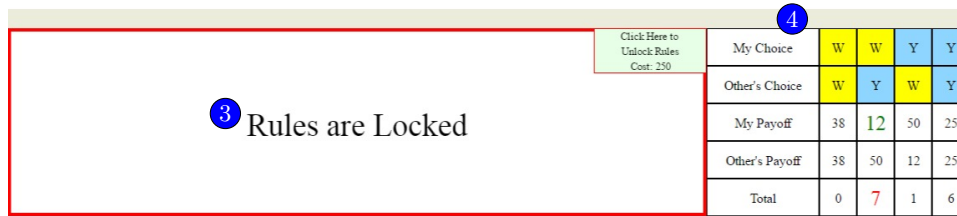
(a) Unlocked Rules View of Constructor Part of the Screen

(b) Locked Rules View of Constructor Part of the Screen

Figure 1: Interface Screenshot. *Notes*: **(a)** Locked rules view of the interface. The screen consists of four main parts: 1) *history* of play at the top of the screen; 2) current *set of rules* in the middle of the screen; 3) rule *constructor* at the bottom/bottom-left of the screen; and 4) *payoff table* at the bottom-right of the screen. No changes can be made to the rule set while rules are locked. To unlock the rule constructor, the subject must press the "Click Here To Unlock Rules" button. The cost to unlock rules is displayed on the button and is added to to the total amount in the top right once the button is clicked. **(b)** Unlocked rules view of the constructor part of the interface. The history, the current rule set, and the payoff table parts of the screen are the same under both views.

Each rule consists of two parts: i) *Input Sequence* - a sequence of action profiles; and ii) *Output Choice* - an action to be played by the subject after the input sequence occurs. Figure 1 displays some examples of rules. For example, Rule #2 (third rule from the left in the middle of Figure 1 (a)) has an input sequence of (W,W), (W,Y), (W,W) and an output of Y. This means that if

the subject plays W, W, and W in the last three periods, and the participant that he is paired with plays W, Y, and W, then this rule will play Y in the next period. The length of the rule is measured by the length of the input sequence. Thus, rule #2 has a length of 3; rule #5 has a length of 2; and rule #6 has a length of 1.

There are several ways that subjects can modify their rule sets during the experiment. First, subjects can use the rule constructor (displayed in Figure 1 (b)) to construct the rules and add them to the rule set. Second, subjects can delete rules from the rule set by pressing the delete button next the rule in the rule list. Third, they can copy an existing rule down to the constructor and create a similar rule by pressing the copy button next to the rule in the rule list. Note that it is not possible to have two rules with the same input sequence but different outputs; if subjects to do this, then they get an error message that says "Conflicting rule in set," and a button that says "Switch Rule" appears. If subjects press this button, it will delete the conflicting rule from the rule set and add the rule from the constructor.

As the match progresses, subjects see the history of play across the top of the screen (marker #1 of Figure 1 (a)). A rule of length n is said to *fit the history* if the input sequence matches the last n periods of the history. For example, if the last period of play in the above history (period 14) is (W,Y), and that sequence is also the input for rule #6 (fifth rule from the left in the middle of Figure 1 (a)), then rule #6 is said to fit the history. If more than one rule fits the history, then the rule with the longest length will determine the choice. If no rules fit the history, then the "default rule" (second rule from the left near marker #2 of Figure 1 (a)) will be selected. The default rule is a memory-0 rule, which is set prior to the start of the first match and can be switched at any time prior to each supergame or when the rule constructor is unlocked. Additionally, since no history is available in the first period, we require subjects to specify the "first-period rule," which selects an action only in the first period. Similar to the default rule, the first-period rule is set prior to the start of the first match and can be switched at any time prior to each supergame and when the rule constructor is unlocked.

2.2   Design Details

Given the relative complexity of the interface, our goal was to make sure that subjects understood the interface, had a sufficient amount of time, and were able to construct a meaningful strategy even before the first supergame began. To this end, we took steps that went beyond the typical experiment. Specifically, upon entering the lab, subjects were randomly assigned to a computer and given a handout containing the instructions. The experiment proceeded as follows: first, recorded video and audio instructions were played on a projector at the front of the lab and also displayed on each computer terminal. Second, subjects played a practice match against a computer that was playing randomly. Third, subjects completed a ten-question quiz to make sure that they understood the interface. To ensure that subjects devoted sufficient effort to following the instructions, the quiz was monetarily incentivized. Fourth, subjects were given ten minutes to construct their initial strategy. Fifth, the experiment began. Next, we describe each of these steps in more detail.

2.2.1 Video Instructions

The video instructions took approximately 20 minutes. The video was displayed on each subject's computer screen, as well as on the screen at the front of the laboratory. The audio was played on the speakers from the computer at the front of the laboratory. The video followed the paper instructions word for word and was read by a computerized voice to ensure a neutral tone. The video provided an extensive explanation of all the steps of the experiment and an elaborate demonstration of the experimental interface. Subjects learned whether editing their rule sets would be costly or costless during these video instructions.

2.2.2 Practice Match

After the instructions, there was a practice match that lasted approximately ten minutes. In the practice match, subjects played against a computer that was playing randomly. Note that no information about the actual payoff table had been provided at this point. Furthermore, the payoff table during the practice match contained all zeros; thus the focus was on being able to construct and modify the rule set and getting comfortable with the interface, rather than on thinking about payoffs and the rules to be used in the experiment.

2.2.3 Quiz

After the practice match, there was a quiz to make sure that everyone understood the instructions. The quiz was incentivized and served two purposes: i) to provide incentives for paying attention to the instructions; and ii) to isolate interaction and data analysis to the participants who understood the interface. Specifically, each participant could earn $10.00 for answering all ten questions correctly. If participants made at least one mistake, they would earn $0.00. Upon making the first mistake, since subjects would not be able to earn the $10.00, their quiz would stop. Subjects were told that they had ten minutes for the quiz, but this constraint was never binding. Subjects learned these details before the instructions began.

To isolate the subjects who understood the interface, we decided on the most stringent criteria and divided the subjects so that the first group consisted of those who received $10.00 for the quiz (subjects who answered all questions correctly). These subjects were told the number of participants in their group and that all participants had received $10.00 for the quiz. Thus, they knew the number of participants in their group and that all participants in the group had an excellent understanding of the interface. The second group consisted of all subjects who received $0.00 for the quiz and, at most, one subject who received $10.00 for the quiz (in case an odd number of subjects received $10.00 for the quiz). Subjects in the second group were told only the number of participants in their group. *We used data only from the first group (74 participants in total) for the analysis carried out in this paper.*[4]

---

[4]A similar experimental interface was also used in Romero and Rosokha (2015), as well as in a pilot for this experiment (details on pilot experiment in Appendix D). In both previous experiments, we used an incentivized quiz in which subjects had ten questions and could earn $0.50 if they answered correctly on their first try, $0.25 if they

2.2.4 Constructing the Initial Set of Rules

After the quiz, subjects were given one minute to focus exclusively on the payoff table to be used in the experiment. Next, subjects were given ten minutes to construct their initial set of rules. During that ten minutes, subjects where presented with a split-screen view of the interface (Figure 2). On the left side of the screen, which we named "hypothetical rules," subjects were able to see the choice that their hypothetical rules would make for a variety of histories. Note that they would have to fill in their own histories for consideration. On the right side of the screen, which we named "Starting Rules," subjects could store rules for the start of the first match. Thus, subjects could consider a multitude of "what-if" scenarios without changing their starting set of rules. In this way, subjects could get comfortable with their rule sets and gain some experience with the game without actually playing.

---

answered correctly on their second try and $0.00 if they didn't get the question right on their first two tries, at which point they were given a hint and still had to correctly answer the question. We found that most subjects were able to correctly answer a large percentage of the questions quickly (in about five minutes). However, several subjects took much longer (up to 20 minutes) to complete the quiz. The subjects that completed the quiz quickly often did the best, while those that took the longest on the quiz often did the worst. This led to several potential problems:

1. Subjects that make a large number of incorrect guesses may not understand the complex interface.

2. Subjects that did well on the quiz had to wait a long time for a small number of subjects to finish.

3. Subjects that are waiting for others to finish may think that everyone else is taking a long time (even if they are waiting for only one subject to finish) and, therefore, believe that no one else understands what is going on.

Based on this previous experience, we designed the quiz with the three corresponding features in an attempt to avoid the above problems:

1. In order to ensure that all subjects understand the interface, we apply the stringent all-or-nothing incentives. We then focus our analysis on those subjects that answered all questions correctly.

2. Since the analysis focuses only on the subjects that correctly answered all quiz questions, once a subject misses a question (and, therefore, his payment has been determined), the quiz ends for him. This ensures that subjects don't have to wait a long time for a subject who doesn't understand to finish the quiz.

3. After all subjects have finished the quiz, they are divided into a high quiz group and a low quiz group. The high quiz group is told the exact number of subjects that are in their group, and that every subject in their group correctly answered all questions on the quiz. This ensures that there is common knowledge in the high quiz group that all subjects in that group understand the interface.

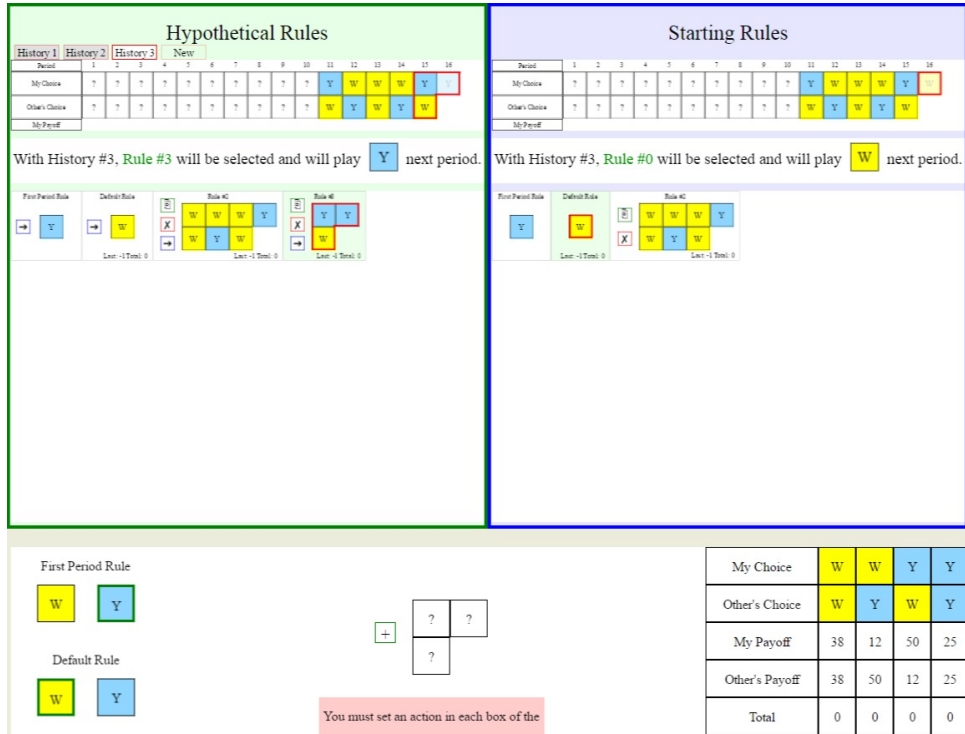This setup gives us confidence that all the subjects analyzed understand the interface.

Figure 2: Split-screen screenshot. *Notes*: To construct the initial set of rules, subjects were presented with a split-screen view of the interface. On the left side of the screen, labeled "Hypothetical Rules," subjects could construct a variety of strategy-history combinations for consideration. On the right side of the screen, labeled "Starting Rules," subjects could maintain a "starting" rule set that they would like to start the experiment with. Subjects could move rules from the left side to the right side using the ➡ button.

Note that an alternative approach —whereby subjects become familiar with the interface through playing against each other via the direct-response (Dal Bó and Fréchette, 2015)— is not the best option for our experiment for two reasons. First, the direct-response with costs creates strategically different incentives for the participants. Specifically, subjects may select their strategies so as to minimize the number of action changes, which directly affects the types of strategies that subjects choose. Second, our motivation is for scenarios in which direct-responding may not be practical or feasible to begin with.

2.2.5 Experiment

The steps described in Sections 2.2.1-2.2.4 took approximately 45 to 50 minutes and were implemented so that the analysis could be carried out starting with the first supergame. After these steps, the experiment began. The experiment consisted of ten supergames. At the beginning of each supergame, participants were randomly paired and remained in the same pairs until the end of the supergame. The number of periods in each supergame was determined using a geometric distribution with a probaiblity of continuation of $\delta = .98$. To ensure a valid comparison, the same

sequence of random numbers was used in every session, resulting in the same number of periods within each supergame (Figure 3 (a)). Figure 3 (b) displays the stage game used in our experiment.

Supergame:	1	2	3	4	5	6	7	8	9	10
Number of Periods:	46	36	52	85	60	7	68	41	44	42

38,38	12,50
50,12	25,25

(a) Supergame Lengths   (b) Stage Game

Figure 3: Supergame Details.

Prior to each supergame (including the first one), subjects had one minute to make changes to their strategies with no costs. Once a supergame began, subjects had one minute per period to make changes to their rules and to confirm their action (they were given a ten-second warning before their time expired in a given period). On the history section of the screen, subjects could see a preview of the action that their current rule set would play in the next period. To confirm an action, subjects had to click the preview of their next-period action. They were required to lock the rule constructor before confirming their action. If a subject failed to confirm the action within a minute, the rule constructor was locked automatically, the action was confirmed automatically, and the experiment proceeded to the next period. Once both participants in a pair confirmed their actions, the game proceeded to the next period. In practice, the one-minute time limit was rarely reached, and supergames proceeded relatively quickly.

By providing subjects the freedom to modify their strategies within each period for a whole minute (if they chose to do so), we gave up some control over the duration of the experiment. This resulted in a situation in which some matches proceeded more slowly than others, and in two of the sessions, we had to stop the experiment before reaching the tenth supergame (one session after supergame 7 and another session after supergame 8).

2.3 Treatments

The experiment consisted of two between-subject treatments: *costless* strategy adjustment ($C = 0$) and *costly* strategy adjustment ($C = 250$). In the costless strategy adjustment treatment, subjects incurred a cost of $C = 0$ for changing their strategy during the supergame. In the costly strategy adjustment treatment, subjects incurred a cost of $C = 250$ for changing their strategy during the supergame.[5] To make a change during the supergame, the subject needed to click the "unlock

---

[5]The cost of editing the rule set was set to $C = 250$, which is five times the highest stage game payoff, 6.5 times the cooperative payoff, and ten times the stage game equilibrium payoff. We chose this based on a pilot experiment that we ran and describe in Appendix D. The chosen cost is high enough that it will have a substantial impact on expected future earnings, but low enough that a subject still can benefit from making a change in expectation. For example, if subject #1 is playing ALLC and subject #2 is playing ALLD, then subject #1 would get an expected benefit of 650 by switching to ALLD, so the cost of editing the rule set is 38% of that benefit. Based on the pilot and these calculations, the chosen cost is high enough to impact the decision to make the adjustment, while not so prohibitively high as to make any adjustment not worthwhile.

rules" button, which would cause the rule constructor to be uncovered (marker #3 in Figure 1(a) would change to marker #3 in Figure 1(b)). The subject could then freely edit his rules until either i) he clicked the "lock rules" button, or ii) the allotted time for the period expired. Subjects incurred the cost every time they clicked the "unlock rules" button and, therefore, could incur multiple costs in the same period. In the $C = 0$ treatment, subjects still had to click the "unlock rules" button but incurred zero cost for doing so.

Treatment	Stage Game Payoff	Continuation Probability	Sessions	N. Sub.	Description
$C = 0$	C D C 38,38 12,50 D 50,12 25,25	$\delta = .98$	3	38	*Costless* strategy adjustment within supergames
$C = 250$		$\delta = .98$	3	36	*Costly* strategy adjustment within supergames

Table 1: Treatments summary

Table 1 presents a summary of the two treatments. To reiterate two points: i) the only difference between the treatments is whether or not strategy adjustment during the supergame is costly; and ii) in both treatments, subjects had a chance to construct and modify strategies at no cost for the ten minutes before the start of the first supergame, and for one minute before each subsequent supergame. Therefore, if subjects make no changes to their strategies during the supergame, then the two treatments should give the same results, as the cost is only incurred for making changes within a supergame.

3   Results

Recall that in order to change the set of rules, the subjects need to click an unlock button. Not surprisingly, we find a significantly higher number of unlocks in the $C = 0$ treatment than in the $C = 250$ treatment (2.37 unlocks per subject per supergame in the $C = 0$ treatment vs. 0.18 unlocks per subject per supergame in the $C = 250$ treatment). The main results of this paper examine the effect that this difference has on subjects' ability to learn and cooperate. In Section 3.1, we present the results on aggregate cooperation observed during our experiment. In Section 3.2, we examine how the unlock cost affects subjects' ability to experiment within and across supergames. In Section 3.3, we examine strategies that subjects use and how they change throughout the experiment.
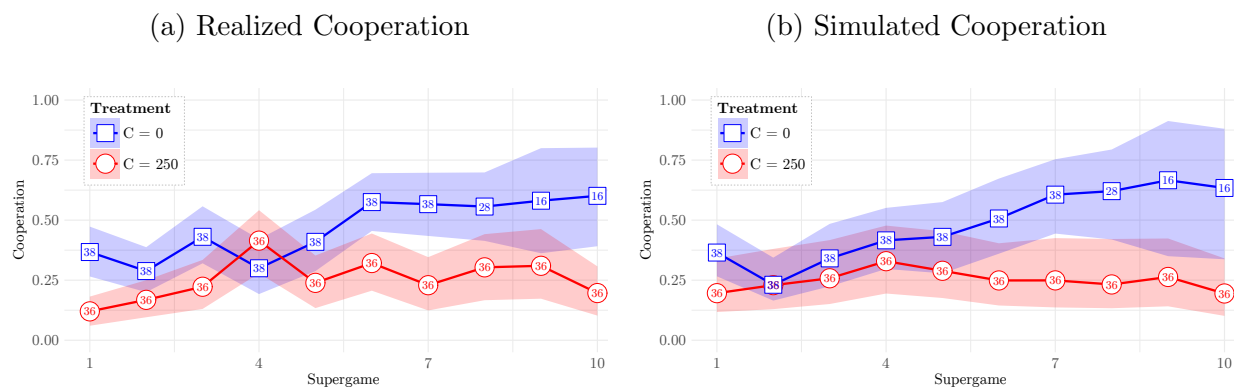
3.1 Cooperation

Figure 4: Cooperation Rate Across Supergames. *Notes*: **(a)** Unit of observation is subject's realized average cooperation over all periods in the supergame. **(b)** Unit of observation is subject's simulated average cooperation over 50 periods. Cooperation is simulated using subjects' rule sets taken at the beginning of each supergame. The simulations are run with all possible matchings, so each subject is matched with every other subject for a total of $n(n-1)$ pairs. The number inside the circle/rectangle represents the number of subjects. 95% bootstrap confidence intervals are superimposed. Supergames 8-10 in the $C = 0$ treatments have fewer observations because some sessions were stopped due to time.

Figure 4 and Table 2 give information about the cooperation rates in the two treatments. The cooperation rate is defined as the percentage of a periods that a subject cooperated over a given range of periods. Figure 4 (a) presents the evolution of cooperation across supergames in our experiment. Each point represents the average cooperation rate among all subjects over the corresponding supergame. Figure 4 (b) presents the evolution of cooperation simulated using subjects' rule sets taken at the beginning of each supergame. Table 2 presents a further breakdown of cooperation rates within the supergame. Cooperation rates are provided for the first period, for the first four periods, for the last four periods, and for all periods for the different treatments during the early stages (supergames 1-5) and the later stages of the experiment (supergames 6-10). In the table, we provide bootstrapped standard errors, as well as results of the probit regressions with standard errors clustered at the session level. Figure 4 and Table 2 yield the following result:

Result 1 *Cooperation is higher when rule set adjustment is costless ($C = 0$) than when rule set adjustment is costly ($C = 250$).*

	Supergames 1-5				Supergames 6-10			
Periods:	First	First 4	Last 4	All	First	First 4	Last 4	All
C = 0	0.553 (0.036)	0.425 (0.025)	0.332 (0.031)	0.359 (0.026)	0.713 (0.038)	0.649 (0.033)	0.528 (0.04)	0.573 (0.035)
	⋛	⋛	⋛	⋛	∨	⋙	⋙	⋙
C = 250	0.522 (0.037)	0.349 (0.026)	0.211 (0.025)	0.233 (0.024)	0.528 (0.037)	0.344 (0.028)	0.257 (0.03)	0.272 (0.029)

Table 2: Average Cooperation. *Notes*: Unit of observation is subject's realized average cooperation rate. Bootstrapped standard errors are in parentheses. Tests are carried out using probit regressions with standard errors clustered at the session level. $>$, $\gg$, and $\ggg$ denote significance at the 0.10, 0.05, and 0.01 levels, respectively.

Result 1, the main result of this paper, states that cooperation is higher when no cost is incurred to adjust the rule set ($C = 0$) than when a cost is incurred to adjust the rule set ($C = 250$). Table 2 shows that the difference in cooperation rates (over the last four periods and all periods) between the $C = 0$ and $C = 250$ treatments is significant at the 0.05 level in the later stages (supergames 6-10). In addition, Figure 4(a) shows that the realized average individual cooperation rate is higher in the $C = 0$ treatment in all supergames, except for supergame 4 and Figure 4(b) shows that simulated average individual cooperation rate is higher in the $C = 0$ treatment in all supergames, except for supergame 2.

We also find that cooperation increases across supergames when $C = 0$ and stays constant when $C = 250$. This can be seen by comparing cooperation in the $C = 0$ treatment over supergames 1-5 (0.359) to that in supergames 6-10 (0.573), which is significant (p-value 0.001) using a matched-pair randomization test. In contrast, in the $C = 250$ treatment, cooperation is not significantly different in supergames 1-5 (0.233) versus supergames 6-10 (0.272, p-value 0.441 using a matched-pair randomization test). This can also be seen in Figure 4, as the $C = 0$ line is trending upward, while the $C = 250$ line is relatively flat.

3.2 Within Supergame Adjustments

To understand what is causing the difference in cooperation between the two treatments, we examine adjustments within the supergame. We find that, even when it is costly, subjects make changes to their rule sets during the supergame, which highlights the importance of adjustments during long repeated interactions. In addition, we find significantly more unlocks and significantly larger payoff changes during the supergame in the $C = 0$ treatment as compared to the $C = 250$ treatment. All of this suggests that more experimentation and easier adjustments to the strategies could be one of the causes of the higher level of cooperation in the $C = 0$ treatment.
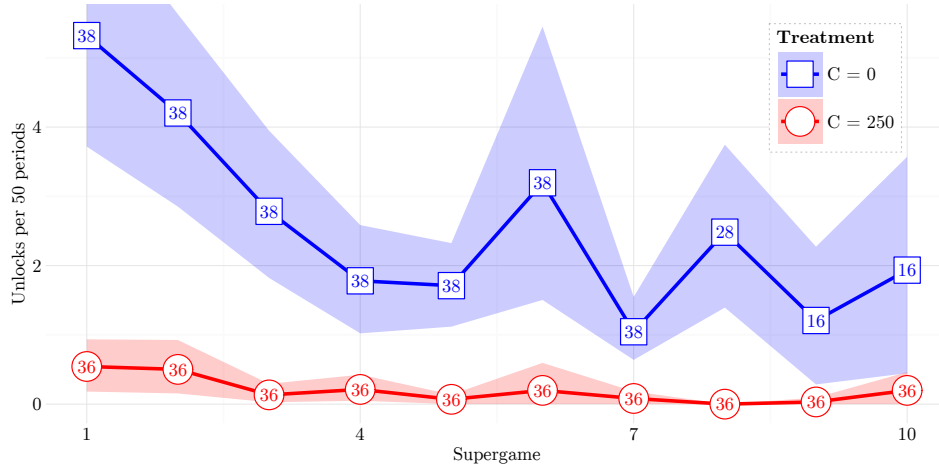
Figure 5: Unlocks Per Period. *Notes*: 95% bootstrap confidence intervals are superimposed. The unit of observation is number of unlocks per 50 periods for each subject-supergame. The number inside circle/rectangle represents the number of units of observation in each supergame.

Figure 5 presents the average number of times per period that subjects click the unlock button for each supergame. The figure shows that the number of times that subjects unlock the rule constructor is significantly higher in the $C = 0$ treatment than in the $C = 250$ treatment in each supergame. In addition, the figure suggests that the number of unlocks is decreasing across supergames in both treatments. More precisely, the average number of unlocks per 50 periods in the first five supergames versus the last five supergames is 3.16 compared to 2.07 in the $C = 0$ treatment (p-value=.010) and 0.29 compared to 0.10 in the $C = 250$ treatment (p-value=.046).[6] Finally, the figure shows that even when unlocking is costly, subjects still choose to unlock the rule constructor within the supergame. (There were 64 unlocks in the $C = 250$ treatment across all supergames.) These results demonstrate that subjects are unlocking the rule constructor while the supergame is progressing, especially in the $C = 0$ treatment.

Though Figure 5 does provide evidence that subjects are unlocking the rule constructor within the supergame, it may not be sufficient evidence that subjects have actually changed their strategies within the supergame. In order to determine whether subjects are changing their strategies within the supergame, we consider the three possible scenarios that could occur after a subject clicks the unlock button:

- **Scenario #1:** A subject unlocked the rule constructor and made an adjustment to his rule set in order to change his underlying strategy.

- **Scenario #2:** A subject unlocked the rule constructor but did not make any adjustments to his rule set, and, therefore, made no change to his underlying strategy.

- **Scenario #3:** A subject unlocked the rule constructor and made an adjustment to the rule

---

[6]*p*-values are obtained using a non-parametric matched-pairs randomization test.

set, but only to implement his underlying strategy. An extreme example of this would be a subject who fully intends to play the Alternator strategy, and does this by unlocking the rules in every period and changing only the default rule. This subject would appear to have many unlocks, while the underlying strategy would stay the same.

In Scenario #1, subjects are changing their strategies, while in scenarios #2 and #3, they are not. Scenario #3 is particularly problematic because it cannot be distinguished from Scenario #1 in the data. Thus, looking solely at the number of unlocks may be overstating the number of actual strategy changes.

Given that the number of unlocks may overstate the number of strategy adjustments within the supergame, we next turn to joint payoffs as additional evidence of adjustments made within the supergame. If subjects use relatively simple strategies (e.g., those that can be represented by one or two state automata) and do not change their strategies within the supergame, then play will converge to a deterministic sequence of action profiles after the first several periods. Therefore, the average joint-payoff will not change from periods 4-15 of the supergame to the last 12 periods of the supergame.[7] Note that this is true even if the two problematic scenarios (Scenarios #2 and #3) stated above occur. Alternatively, if subjects use relatively simple strategies, but they change their strategies within the supergame, then the average joint-payoff may be different in periods 4-15 of the supergame when compared to the last 12 periods of the supergame. For example, consider a supergame of length 50 in which player 1 plays the TFT strategy, and player 2 starts by playing the TFT strategy for the first 20 periods, but then switches to the ALLD strategy for the remaining periods. In this case, the average payoff will be 38 for each player in periods 4-15 but only 25 for each player in the last 12 periods of the supergame.

Figure 6 examines the change in average payoffs from the beginning to the end of the supergame. Panel (a) shows the change in joint average payoffs from periods 4-15 to the last 12 periods of each supergame. The length of the arrow indicates the magnitude of the change in payoff from the beginning to the end of the supergame. A green arrow denotes a pair that had a Pareto improvement. Pairs that had no change are denoted with a circle (the relative size of the circle captures the number of pairs at that point). Panel (b) compares the magnitude of these changes between the two treatments.

---

[7]Previous literature on indefinitely repeated prisoner's dilemma games has suggested that a majority of players use memory-1 strategies (Dal Bó and Fréchette, 2016). If both players use memory-1 strategies, then play will converge to a deterministic pattern of action profiles of length 1, 2, 3 or 4 by, at the latest, the fourth period of the supergame. In addition, a sequence with a pattern of length 1, 2, 3 or 4 will always give the same average payoff over a block of 12 periods. Therefore, we compare average payoff in periods 4-15 with the average payoff in the last 12 periods of the supergame.

(a) Change in the Average Joint Payoffs
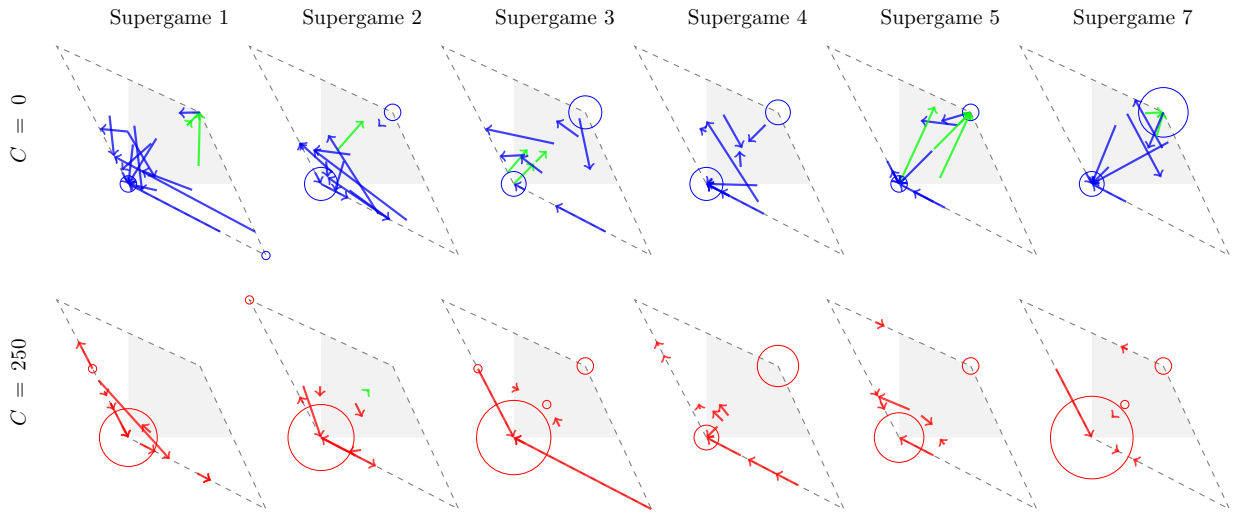
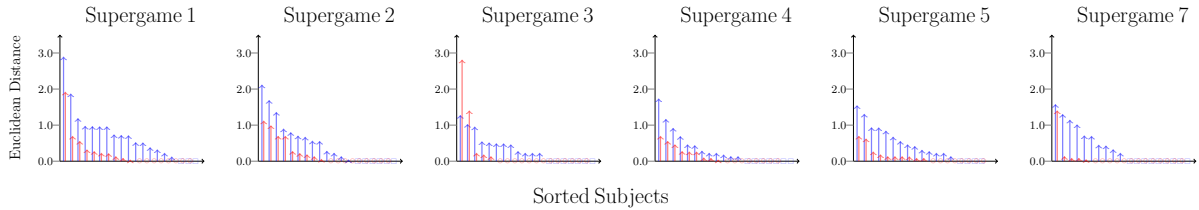(b) Change in the Average Joint Payoffs Sorted by Magnitude

Figure 6: Payoff Changes Within Supergames. *Notes*: **(a)** The arrows denote the change from the average joint payoff in periods 4-15 to the last 12 periods. Green arrows highlight those pairs that had strict increases in both payoffs from periods 4-15 to the last 12 periods of a given supergame. The circles represent pairs that had no change in payoff, with the relative size indicating the number of subjects at a given point. **(b)** Cumulative distribution of the change magnitude from the average joint payoffs. $C = 0$ is in blue. $C = 250$ is in red. Supergame 6 is omitted because it had only seven periods and, therefore, trivially would lead to no change in payoff. All other supergames contained at least 27 periods, which ensured that there was no overlap between periods 4-15 and the last 12 periods of the supergame.

Figure 6 provides evidence that subjects in the $C = 0$ treatment change their strategies more than the subjects in the $C = 250$ treatment. This is evidenced by the significantly larger number of arrows with positive length in each supergame of the $C = 0$ treatment (16, 13, 12, 12, 14, 10, respectively, or 68% of all pairs) when compared to the $C = 250$ treatment (10, 9, 5, 9, 10, 5, respectively, or 45% of all pairs, p-value < 0.001, using a non-parametric permutation test). In addition to the higher number of arrows, the bottom panel shows that the arrows also have a greater magnitude in the $C = 0$ treatment than in the $C = 250$ treatment. To confirm, we find that the average magnitude in the $C = 0$ treatment (0.48) is significantly higher than the average

16

magnitude in the $C = 250$ treatment (0.20, p-value $< .001$), using a non-parametric permutation test. These findings are summarized in Result 2.

Result 2 *Subjects make changes to their strategies within the supergame in both treatments. Significantly more of these changes occur when rule set adjustment is costless ($C = 0$) than when rule set adjustment is costly ($C = 250$).*

Result #2 states that subjects make substantial changes to their strategies within an ongoing interaction, and significantly more changes occur in the $C = 0$ treatment. In the $C = 250$ treatment, despite the cost of unlocking the rule set—five times the highest stage game payoff—subjects still unlocked their rule sets 64 times in that treatment. The fact that subjects were willing to incur this substantial cost to change their rule set suggests that changing the rule set during the interaction was important to them. In the $C = 0$ treatment, we find significantly more changes to the strategies than in the $C = 250$ treatment. This is shown by the significantly higher number of times that subjects unlocked their rule sets, as well as significantly more and significantly larger changes in joint payoffs from the beginning to the end of the supergame.

3.3 Discussion of Strategies

Although the interface provides a unique way to elicit strategies, this experiment is not designed to investigate the strategies. First, we cannot necessarily rely on the constructed strategies in the $C = 0$ treatment, as there is no explicit incentive for subjects to construct strategies. Second, the evidence that subjects modify their strategies within a supergame means that the estimation results should be interpreted with caution. Despite these caveats, we still may be able to get some idea of the strategies that people play by looking at rules and estimating strategies. In Section 3.3.1, we provide data on the number and type of rules constructed in the two treatments, and in Section 3.3.2, we estimate strategies from the data using the maximum likelihood approach.

3.3.1 Rules In Set

Figure 7 presents the total number of rules across supergames (panel a) and the number of rules by memory length (panel b). We find no difference between the treatments in either the number or the length of rules.

17

(a) Number of Rules Across Supergames.

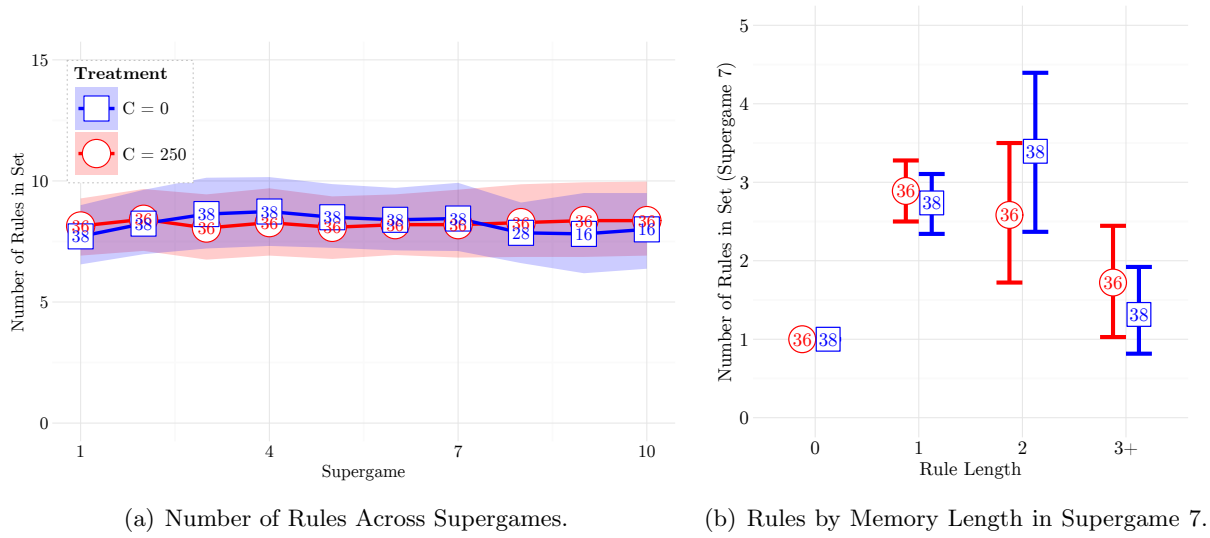(b) Rules by Memory Length in Supergame 7.

Figure 7: Number of Rules in Set.

Next, we take a more in-depth look at which rules are constructed in each treatment. Figure 8 presents the proportion of subjects having the memory-0 and memory-1 rules at the beginning of each supergame.
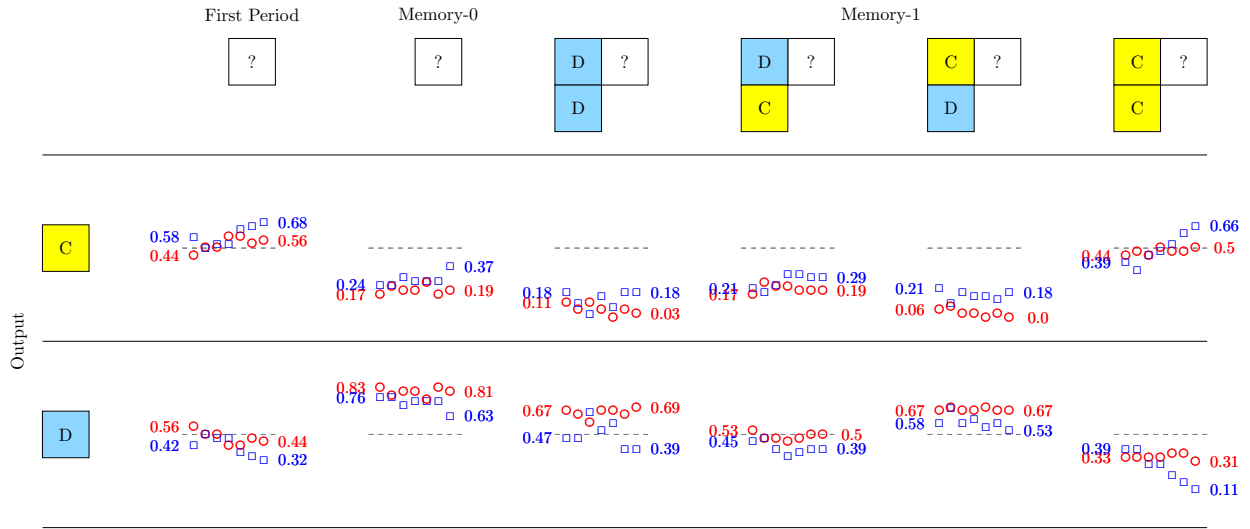

Figure 8: Proportion of Memory-1 Rules. *Notes*: The proportion of each of the memory-0 and memory-1 rules at the beginning of each supergame. Blue squares denote $C = 0$ treatment. Red circles denote $C = 250$ treatment. Dashed line denotes 50%.

The main difference between the two treatments is that the proportion of $CC \to C$ and $\to C$ rules increases across supergames in the $C = 0$ treatment (0.39 to 0.66 and 0.24 to 0.37, respectively), while the proportion of corresponding rules in the $C = 250$ treatment stays largely the

same (0.44 to 0.5 and 0.17 to 0.19, respectively). However, looking at rules by themselves may not be sufficient because multiple combinations of rules can comprise the same strategy (for example, subjects can make TFT with either {C(first), $\to C$, $DD \to D$, $CD \to D$} or {C(first), $\to D$, $CC \to C$, $DC \to C$}). Next, to get a better idea about the complete strategies that subjects play, we use the strategy frequency estimation method (Dal Bó and Fréchette, 2011). Note that we implement a method that infers strategies from observed actions rather than from constructed rules (as done in Romero and Rosokha (2017)), because there are no explicit incentives to construct a strategy with rules in the $C = 0$ treatment.

3.3.2 Strategies

In this section, we use the maximum likelihood estimation approach (Dal Bó and Fréchette, 2011) to find the most likely strategies among the population, given the observed action choices. As noted earlier, all sessions contain at least seven supergames, but some of the sessions were stopped early due to time constraints. Therefore, for the analysis, we partitioned the data into two sets of three supergames: early stages (supergames 1-3) and late stages (supergames 5-7).[8]

Table 3 presents frequencies estimated for the full 20-strategy set from Fudenberg, Rand, and Dreber (2012). Bootstrapped standard errors are calculated by drawing 100 random samples (with replacement) of the appropriate size, estimating the maximum likelihood estimates of the strategy frequencies corresponding to each of these samples, and then calculating the standard deviation of the sampling distribution (Efron and Tibshirani, 1986).

Supergames	Cost	ALLD	Tit For Tat	Grim	False Cooperator	Exp. Tit For Tat	Tit For 2 Tats	Lenient Grim 2	Tit For 3 Tats	Lenient Grim 3	ALLC	2 Tits for 2 Tats	Exp. Grim 3	β	Cooperation	# of Subjects
1-3	0	0.26*** (0.08)	0.14** (0.07)	0.16** (0.07)	0.08* (0.05)	0.18** (0.08)	0.03 (0.03)	0.08* (0.05)		0.05 (0.05)				0.84	0.38	38
	250	0.45*** (0.10)	0.15** (0.07)	0.14** (0.08)	0.10** (0.05)	0.03 (0.03)			0.06* (0.04)		0.03 (0.03)	0.02 (0.03)		0.90	0.20	36
5-7	0	0.18*** (0.06)	0.36*** (0.09)		0.03 (0.03)	0.04 (0.04)	0.13* (0.08)	0.04 (0.04)	0.06 (0.05)		0.07* (0.05)		0.05* (0.03)	0.91	0.53	38
	250	0.39*** (0.08)	0.17* (0.10)	0.17* (0.12)	0.10** (0.05)					0.06* (0.05)		0.07* (0.05)		0.94	0.26	36

Table 3: MLE Strategy Estimates. *Notes*: Bootstrapped standard errors are in parentheses. *, **, ***, denote significance at the 0.10, 0.05, and 0.01 levels, respectively.

We find that subjects relied predominantly on memory-1 strategies (AllD, TFT, Grim, False Cooperator, ETFT, ALLC). For supergames 1-3, the cumulative proportions of memory-1 strategies are 0.82 and 0.90 for the $C = 0$ and $C = 250$ treatments, respectively. For supergames 5-7, the cumulative proportions of memory-1 strategies are 0.68 and 0.83 for the $C = 0$ and $C = 250$

---

[8]We chose this partition because it is the partition with the largest number of supergames and no overlap, given our seven supergames of complete data.

treatments, respectively. Furthermore, our results are consistent with Dal Bó and Fréchette (2016), who review 17 separate treatments (across six studies) and find that in 15 of them the majority of subjects use one of the three simple commonly observed strategies (ALLD, TFT, GRIM).

The main observation from the rules and the MLE estimates is summarized in the result below:

Result 3 *Subjects develop cooperative strategies when adjusting the rule set is costless ($C = 0$) but not when adjusting the rule set is costly ($C = 250$).*

Result #3 states that the evolution of strategies is different in the two treatments. Though subjects play less-cooperative strategies in the early stages of both treatments, more-cooperative strategies arise when it is costless to adjust strategies ($C = 0$), and less-cooperative strategies prevail when it is costly to adjust strategies ($C = 250$). More specifically, the most common strategies in supergames 1-3 of the $C = 0$ treatment are less cooperative (AllD 0.26, Grim 0.16, False Cooperator 0.08 and Exploitative Tit For Tat 0.18), while more-cooperative strategies are less prevalent (Tit For Tat 0.14, Lenient Grim 2 0.08 and Tit For 2 Tats 0.03). In supergames 5-7 of the $C = 0$ treatment, subjects shift away from the less-cooperative strategies (AllD 0.18, Grim 0.00, False Cooperator 0.03, and Exploitative Tit For Tat 0.04) and towards more-cooperative strategies (Tit For Tat 0.36, Tit For 2 Tats 0.13, and ALLC 0.07). Meanwhile, in the $C = 250$ treatment, the three most common strategies are the same between supergames 1-3 (ALLD 0.45, Grim 0.14, False Cooperator 0.10) and supergames 5-7 (ALLD 0.39, Grim 0.17, False Cooperator 0.10).

To better show this learning, Figure 9 displays MLE estimates for groups of three supergames over time. For expositional purposes, we classified strategies into six types: AllDs, ETFTs, Grims, TFTs, AllC and Alternator. The leftmost entries in the figure correspond to supergames 1-3 in Table 3, and the entries labeled 5-7 correspond to supergames 5-7 in Table 3. Panel (a) shows the evolution of strategy types for the $C = 0$ treatment. In the $C = 0$ treatment, the fraction of TFTs steadily increases (from 0.17 to 0.62), while AllDs and Grims steadily decrease (from 0.38 to 0.22 and from 0.27 to 0.04, respectively). Panel (b) shows the evolution of strategy types for the $C = 250$ treatment. In the $C = 250$ treatment, the fraction of AllDs remains roughly at 0.50 throughout, while the fractions of Grims and TFTs take up the other 0.50.
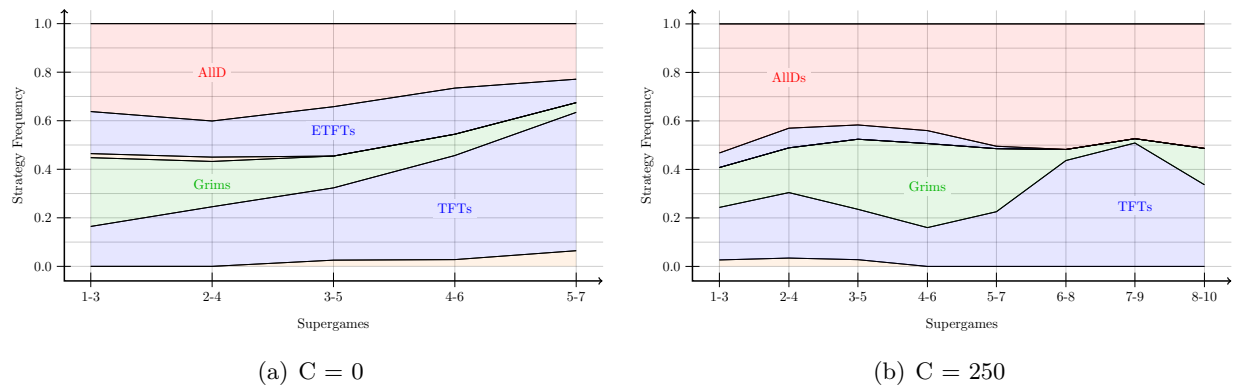
(a) C = 0   (b) C = 250

Figure 9: Evolution of Strategies. *Notes*: MLE estimation is carried out for each of the supergame sets: 1-3, 2-4, ... , 8-10. ALLDs include ALLD and False Cooperator; ETFTs include Exp. Tit For Tat and Exp. Tit for 2 Tats; Grims include Grim, Lenient Grim 2, Lenient Grim 3; TFTs include Tit For Tat, Tit for 2 Tats, Tit for 3 Tats, 2 Tits for 1 tat, WSLS, WSLS 2 Puns.

Strategy estimates corroborate Result #1 in both cooperation levels and cooperation dynamics throughout the experiment. Specifically, consistent with the difference in cooperation rates, subjects in the $C = 0$ treatment use more-cooperative strategies, such as Tit For Tat, more frequently, and less-cooperative strategies, such as ALLD, less frequently than in the $C = 250$ treatment. Additionally, consistent with the evolution of cooperation, subjects in the $C = 0$ treatment learn to play more-cooperative strategies towards the second half of the experiment.

4 Simulations

In this section, we conduct a simulation exercise to provide further intuition about the role of adjustment costs in the learning process. In particular, we use the Genetic Algorithm approach (Holland, 1975) to model the strategy selection and adjustment process. We chose this approach for several reasons. First, the design of the interface fits well with the way that chromosomes typically are encoded with the genetic algorithm. Second, genetic algorithms have been successfully applied to model learning in many economic environments (e.g., Andreoni and Miller, 1995; Bullard and Duffy, 1999). Third, genetic algorithms have been shown to be a good match for behavioral regularities in human subject experiments (e.g., Arifovic, 1994; Duffy, 2006). Though the model does not fit the experimental environment exactly, the simplified environment still yields some intuition about the effect of costs on learning. In particular, higher costs lead to less experimentation, which hinders subjects' ability to learn to cooperate.

4.1 Genetic Algorithm Details

A chromosome, a binary string of length l, represents a strategy. We relate each chromosome to a strategy in our experiment as follows: each bit of a chromosome represents an output of a single

rule.[9] Given that subjects rely primarily on memory-1 strategies, we limit our analysis in this paper to memory-1 strategies. An additional benefit of this simplifying assumption is that computational time declines substantially. Thus, we set the number of bits in a chromosome to be $l = 5$.

We set each bit to represent an output for one of five possible inputs: the first bit is the output of the first-period rule; the second bit is the output of a rule with an input of CC; the third bit is the output of a rule with an input of CD; the fourth bit is the output of a rule with an input of DC; and the fifth bit is the output of a rule with an input of DD. Figure 10 presents an example of three chromosomes that correspond to three of the frequent strategies in our experiment.
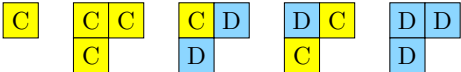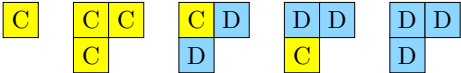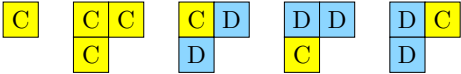
Chromosome	Rule Set					Strategy
0 0 1 0 1	C	CC→C	CD→D	DC→C	DD→D	TFT
0 0 1 1 1	C	CC→C	CD→D	DC→D	DD→D	Grim
0 0 1 1 0	C	CC→C	CD→D	DC→D	DD→C	WSLS

Figure 10: Examples Of Chromosomes Representing Memory-1 Strategies *Notes*: The first bit is the output of the first-period rule; the second bit is the output of a rule with an input of CC; the third bit is the output of a rule with an input of CD; the fourth bit is the output of a rule with an input of DC; and the fifth bit is the output of a rule with an input of DD.

A population of chromosomes, A_t, represents a collection of strategies that a subject considers in generation t. The population of chromosomes evolves over time using three genetic operators – reproduction, mutation, and crossover. Reproduction makes copies of a chromosome i with probability p_i (defined below). Mutation randomly changes the value of each bit within a chromosome with probability p_{mut}. Crossover exchanges the parts of k pairs of randomly selected chromosomes.

The subject chooses one strategy from the set A_t in each generation. In the subsequent generation, if a subject wants to switch to another strategy, then he would incur a cost C. This adjustment cost influences the evolutionary learning process through the reproduction operator. Specifically, the probability of choosing a chromosome i will depend on two parts: i) the fitness, f_i, which measures the average payoff against all other chromosomes in A_t:

$$f_i = \frac{1}{|A_t| - 1} \sum_{j \in A_t / \{i\}} \pi_i \left(i, j \right),$$

---

[9]A more general approach would model both the input and the output of a rule using the genetic algorithm. We leave this for future research.

where $\pi_i(i, j)$ is the total payoff over 50 periods for strategy i when it is matched with strategy j; and ii) the cost, c_i, associated with selecting that chromosome:

$$c_i = \begin{cases} 0 & \text{if } i \text{ was chosen last generation} \\ -C & \text{if } i \text{ was not chosen last generation} \end{cases}$$

We assume the logit process, such that the probability that a chromosome $i \in A_t$ is selected to the next generation is

$$p_i = \frac{e^{\lambda * (f_i + c_i)}}{\sum\limits_{j \in A_t} e^{\lambda * (f_j + c_j)}},$$

where λ determines how stringent the selection is from generation t to generation $t+1$. Figure 11 presents the summary details of the algorithm.

Algorithm 1 Incorporating Costs into Genetic Algorithm

1: Initialize A_0 of n chromosomes and randomly label one as "chosen"
2: For $t = 0$ to T do
3: For all $i \in A_t$ do
4:    Evaluate fitness, f_i
5:    Determine cost, c_i
6: Select n chromosomes to A_{t+1} // Reproduction
7:    Probability that $i \in A_t$ is selected to A_{t+1} is p_i
8:    Chromosome with highest $(f_i + c_i)$ is labeled as "chosen"
9: Mutate each bit of every $i \in A_{t+1}$ with probability p_{mut} // Mutation
10: Exchange parts of k randomly selected chromosomes // Crossover
11: $t = t + 1$

Figure 11: Genetic Algorithm Summary.

4.2 Simulation Results

Figure 12 presents the evolution of cooperation for several combinations of λ and p_{mut} and the two values of C used in our experiment. For expositional purposes, we limit the simulations to the case in which $k = 0$. We present additional simulation results for $k > 0$ in the Appendix E.
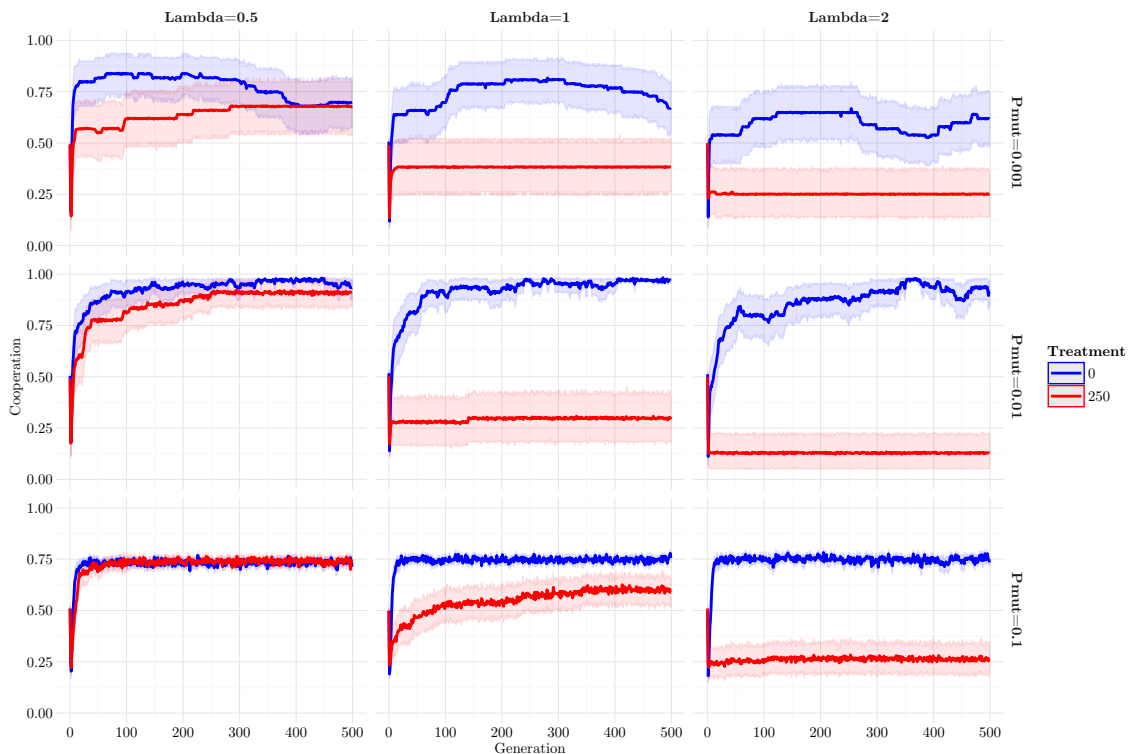
Figure 12: Evolution of Cooperation within the Population. *Notes*: Fifty simulations of population of $n = 50$ chromosomes are conducted. 95% bootstrapped confidence intervals around the average cooperation are superimposed.

We find that, regardless of the costs, cooperation starts out at 50% and immediately dips. Thus, defection is discovered early on. The difference arises as the simulations progress. When costs for strategy adjustment are low ($C = 0$), cooperation is discovered after the defection, while, when costs for strategy adjustment are high ($C = 250$), cooperation is either not discovered or discovered much later.

The simulation results in Figure 12 show three variables that affect cooperation: costs, sensitivity parameter λ, and mutation probability p_{mut}. These variables affect cooperation by varying the agents' ability to experiment with different strategies. Specifically, high costs make experimentation among the strategies in A_t costly, and low costs make experimentation cheap. Similarly, high values of λ lead to less experimentation (via the logistic choice function) and low values of λ lead to more experimentation. Finally, high values of p_{mut} lead to more mutations, which allow for more variation in the set A_t. The simulations show that as the value of λ decreases, there is more cooperation when the costs are $C = 250$. However, as λ continues to decrease beyond $\lambda = 0.5$, the choices become more random, and would eventually lead to less cooperation. So, although experimentation can lead to more cooperation, too much experimentation can be detrimental.

Figure 13 presents the evolution of strategies for the intermediate values of λ and p_{mut}.
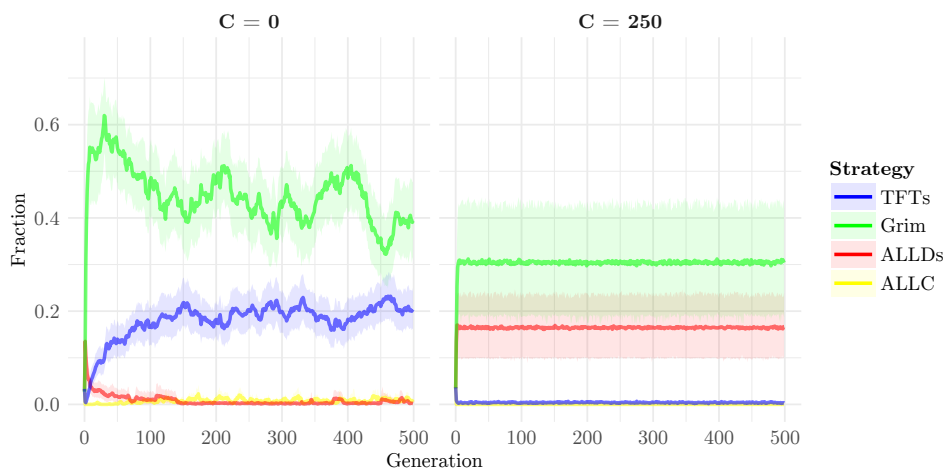
Figure 13: Evolution of Strategies within the Population. *Notes*: Fifty simulations of population of $n = 50$ chromosomes are conducted with $\lambda = 1$, $p_{mut} = .01$, and $k = 0$. 95% bootstrapped confidence intervals around the average fractions are superimposed.

Results in Figure 13 provide further intuition behind our experimental results. Specifically, when costs are high, agents are less likely to change the strategy because experimentation is costly, and, therefore, defective and unforgiving strategies dominate once chosen initially (ALLD and GRIM). When costs are low, even though agents initially choose defective and unforgiving strategies (ALLD and GRIM), experimentation allows them to learn a forgiving strategy (TFT) and to cooperate over time.

5 Conclusion

We report results obtained from the indefinitely repeated prisoner's dilemma with costly strategy adjustment. We find that cooperation is lower when strategy adjustments are costly ($C = 250$) than when strategy adjustments are costless ($C = 0$). This difference occurs even though subjects have the ability to specify the full strategy before the start of the supergame for free. These findings suggest that the ability to learn and experiment during the supergame is an important factor in learning to cooperate. Therefore, in environments in which actions are made through strategies (either explicitly or implicitly), the level of cooperation may depend on how costly it is to modify the strategy.

We identify the source of the difference in cooperation between the two treatments: in the $C = 250$ treatment, less-cooperative strategies predominate, while in the $C = 0$ treatment, subjects learn to play more-cooperative strategies over time. To gain further insight into the mechanism at play, we conducted simulations using a popular evolutionary algorithm. We incorporated the adjustment costs into the algorithm and found that when costs are high, evolutionary agents are less likely to discover the cooperative behavior due to the fact that experimentation is more costly.

The only difference between the two treatments in our experiment is the cost for adjusting

25

strategies during the supergame. Therefore, if players learn across but not within supergames, then the cost should have no effect on cooperation, as all adjustments would be made in the free period between supergames. However, since we see a difference in cooperation between the two treatments, this provides evidence that there is within-supergame learning.

Our results suggest two interesting avenues for future research. The first is to develop new models of within-game learning. These models would likely need to incorporate learning over strategies rather than learning over actions, as standard action-learning models have a difficult time explaining the high incidence of cooperation in repeated prisoner's dilemma games (e.g., Hanaki, Sethi, Erev, and Peterhansl, 2005; Ioannou and Romero, 2014). The second avenue involves developing new techniques of identifying strategies in the presence of within game-learning. These would include either strategy elicitation (e.g. Dal Bó and Fréchette, 2015; Romero and Rosokha, 2017) or new maximum likelihood estimation methods that factor in within-game learning.

References

ANDREONI, J., AND J. H. MILLER (1995): "Auctions with artificial adaptive agents," *Games and economic behavior*, 10(1), 39–64.

ARIFOVIC, J. (1994): "Genetic algorithm learning and the cobweb model," *Journal of Economic dynamics and Control*, 18(1), 3–28.

BIGONI, M., M. CASARI, A. SKRZYPACZ, AND G. SPAGNOLO (2015): "Time horizon and cooperation in continuous time," *Econometrica*, 83(2), 587–616.

BRANDTS, J., AND G. CHARNESS (2011): "The strategy versus the direct-response method: a first survey of experimental comparisons," *Experimental Economics*, 14(3), 375–398.

BREITMOSER, Y. (2015): "Cooperation, but no reciprocity: Individual strategies in the repeated Prisoner's Dilemma," *The American Economic Review*, 105(9), 2882–2910.

BRUTTEL, L., AND U. KAMECKE (2012): "Infinity in the lab. How do people play repeated games?," *Theory and Decision*, 72(2), 205–219.

BULLARD, J., AND J. DUFFY (1999): "Using genetic algorithms to model the evolution of heterogeneous beliefs," *Computational Economics*, 13(1), 41–60.

CAMERA, G., AND M. CASARI (2009): "Cooperation among strangers under the shadow of the future," *The American Economic Review*, 99(3), 979–1005.

CASARI, M., AND T. N. CASON (2009): "The strategy method lowers measured trustworthy behavior," *Economics Letters*, 103(3), 157–159.

CASON, T. N., S.-H. P. LAU, AND V.-L. MUI (2016): "Prior Interaction, Identity, and Cooperation in the Inter-Group Prisoners Dilemma," *Purdue Working Paper*.

DAL BÓ, P. (2005): "Cooperation under the shadow of the future: experimental evidence from infinitely repeated games," *American Economic Review*, pp. 1591–1604.

DAL BÓ, P., AND G. R. FRÉCHETTE (2011): "The evolution of cooperation in infinitely repeated games: Experimental evidence," *The American Economic Review*, 101(1), 411–429.

——— (2015): "Strategy choice in the infinitely repeated prisoners dilemma," *Available at SSRN 2292390*.

——— (2016): "On the determinants of cooperation in infinitely repeated games: A survey," *Journal of Economic Literature*.

DUFFY, J. (2006): "Agent-based models and human subject experiments," *Handbook of computational economics*, 2, 949–1011.

DUFFY, J., AND J. OCHS (2009): "Cooperative behavior and the frequency of social interaction," *Games and Economic Behavior*, 66(2), 785–812.

EFRON, B., AND R. TIBSHIRANI (1986): "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy," *Statistical science*, pp. 54–75.

EMBREY, M., G. R. FRÉCHETTE, AND S. YUKSEL (2015): "Cooperation in the finitely repeated prisoner's dilemma," *Working Paper*.

EMBREY, M., F. MENGEL, AND R. PEETERS (2016a): "Eliciting Strategies in Indefinitely Repeated Games of Strategic Substitutes and Complements," *Working Paper*.

——— (2016b): "Strategy Revision Opportunities and Collusion," *Working Paper*.

ENGLE-WARNICK, J., AND R. L. SLONIM (2004): "The evolution of strategies in a repeated trust game," *Journal of Economic Behavior & Organization*, 55(4), 553–573.

——— (2006): "Learning to trust in indefinitely repeated games," *Games and Economic Behavior*, 54(1), 95–114.

FRIEDMAN, D., AND R. OPREA (2012): "A continuous dilemma," *The American Economic Review*, 102(1), 337–363.

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(2), 720–49.

GREINER, B. (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 1(1), 114–125.

HANAKI, N., R. SETHI, I. EREV, AND A. PETERHANSL (2005): "Learning strategies," *Journal of Economic Behavior & Organization*, 56(4), 523–542.

HO, T.-H. (2008): "Individual learning in games," *The New Palgrave Dictionary of Economics*, 4, 230–235.

HOLLAND, J. H. (1975): *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence.* U Michigan Press.

HONHON, D., AND K. B. HYNDMAN (2015): "Flexibility and Reputation in Repeated Prisoners' Dilemma Games," *Available at SSRN 2562727*.

IOANNOU, C. A., AND J. ROMERO (2014): "A generalized approach to belief learning in repeated games," *Games and Economic Behavior*, 87, 178–203.

LUGOVSKYY, V., D. PUZZELLO, A. SORENSEN, J. WALKER, AND A. WILLIAMS (2017): "An experimental study of finitely and infinitely repeated linear public goods games," *Games and Economic Behavior*.

PEYSAKHOVICH, A., AND D. G. RAND (2013): "Habits of virtue: creating norms of cooperation and defection in the laboratory," *Available at SSRN 2294242*.

ROMERO, J., AND Y. ROSOKHA (2015): "Constructing Strategies in the Indefinitely Repeated Prisoner's Dilemma," *Purdue Working Paper*.

——— (2017): "Constructing Strategies in the Indefinitely Repeated Prisoner's Dilemma," *Purdue Working Paper*.

SHUGAN, S. M. (1980): "The cost of thinking," *Journal of consumer Research*, 7(2), 99–111.

SIMON, H. A. (1957): *Models of man; social and rational.* Wiley.

STAHL, D. O. (2011): "Cooperation in the sporadically repeated prisoners' dilemma via reputation mechanisms," *Journal of Evolutionary Economics*, 21(4), 687–702.

——— (2013): "An experimental test of the efficacy of a simple reputation mechanism to solve social dilemmas," *Journal of Economic Behavior & Organization*, 94, 116–124.

Appendices

A   Quiz

The quiz consisted of ten questions displayed on the screen after the instructions were completed. There were three different types of questions:

1. Given the current rule set, what action will be played next? (Four questions: #1, #4, #7 and #10.)

2. Given the current rule set, what payoffs will be received in the next period if the other subject chooses a given action? (Three questions: #2, #5 and #8.)

3. Modify the rule set so that it plays a given action. (Three questions: #3 add rule, #6 delete rule, #9 swith rule.)

The full screenshot of each of the ten questions is available here: `http://web.ics.purdue.edu/~yrosokha/docs/quizCosts.pdf`

B   Additional Figures and Tables
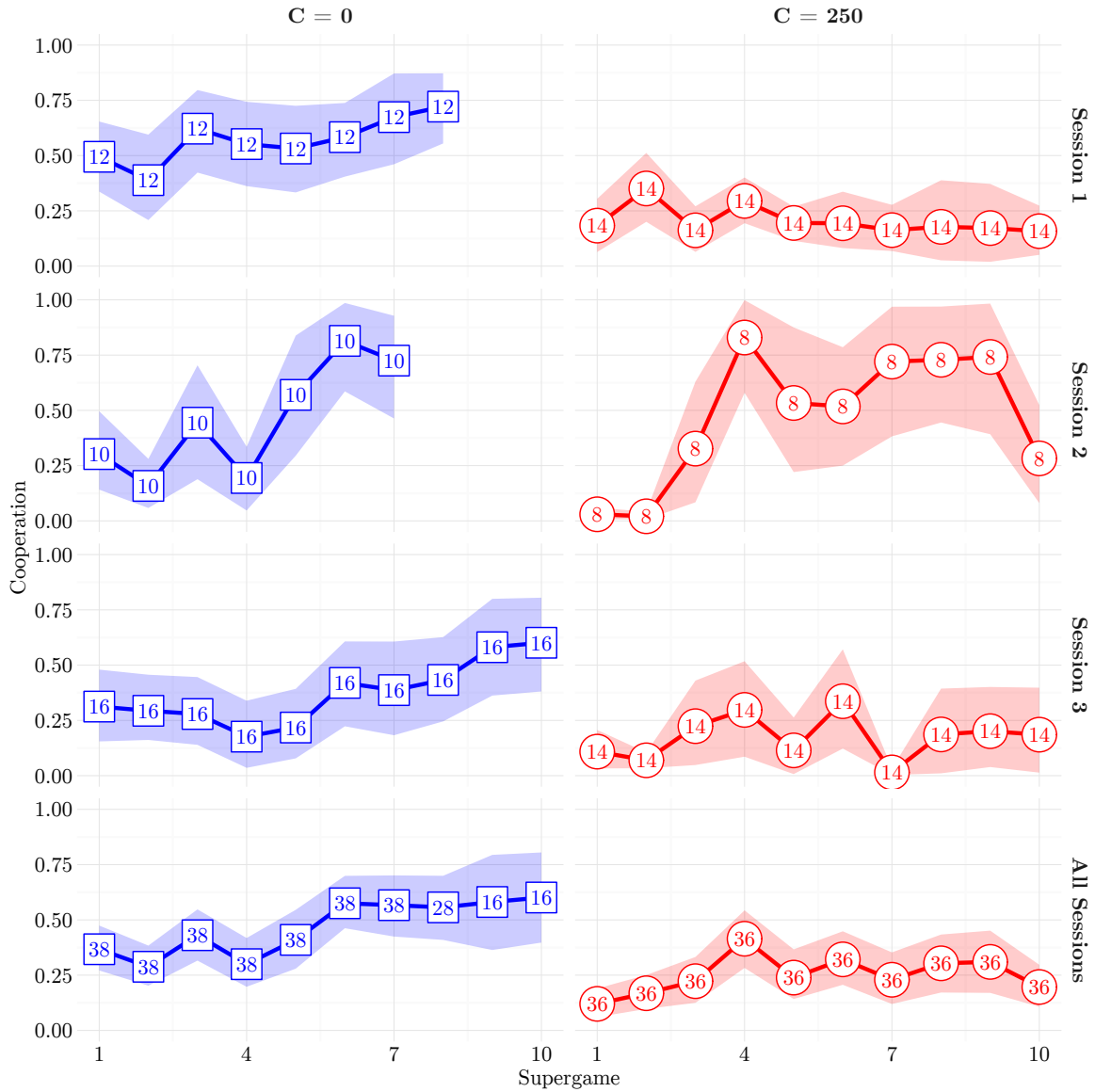
B.1   Cooperation by Session

Figure B.1: Evolution of Cooperation by Session. *Notes*: Cooperation is increasing in all three sessions of $C = 0$. Cooperation is flat in two out of three in $C = 250$ and increasing, then decreasing, in one.

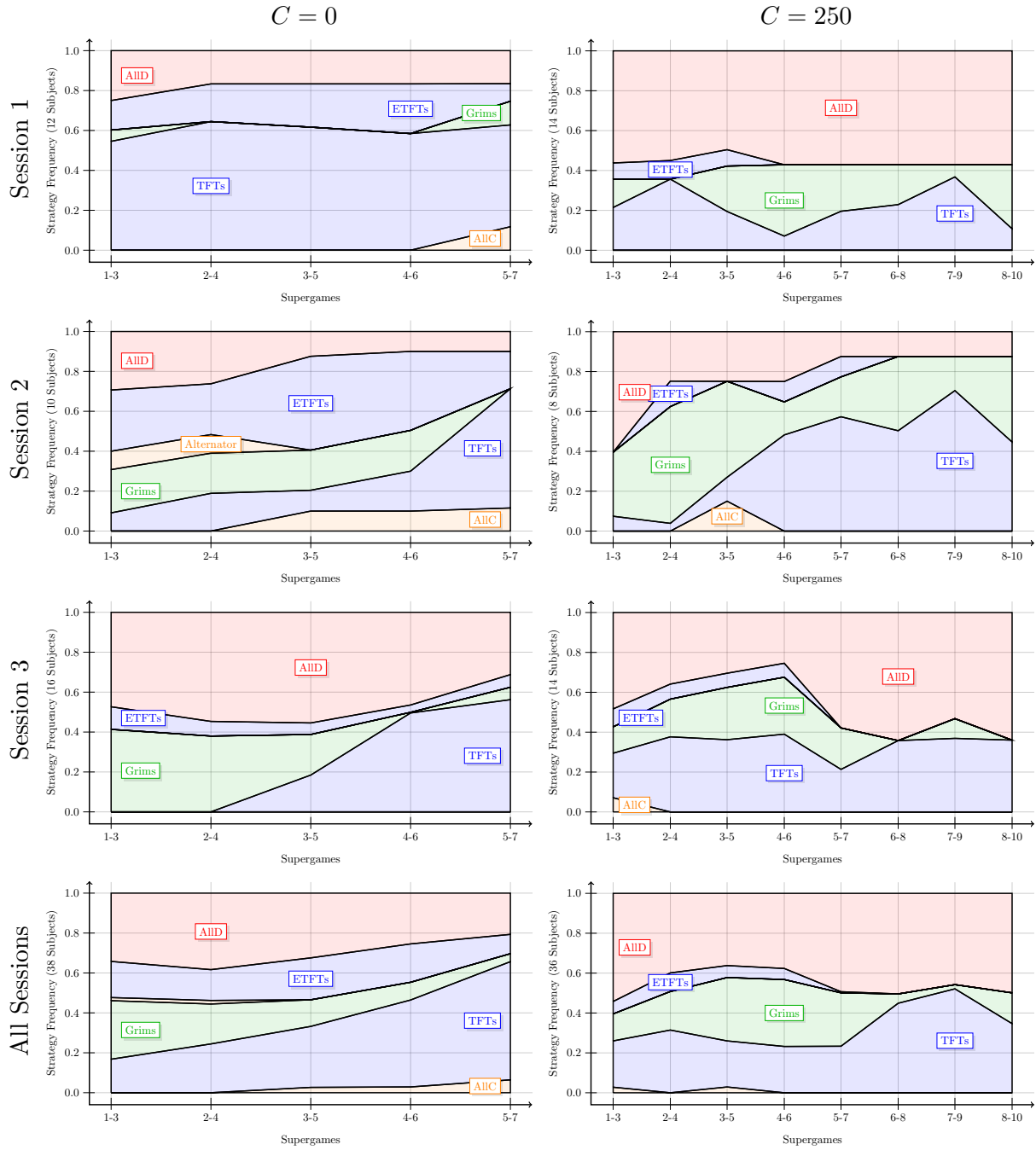## B.2 Evolution of Strategies by Session

Figure B.2: Evolution of Strategies by Session.
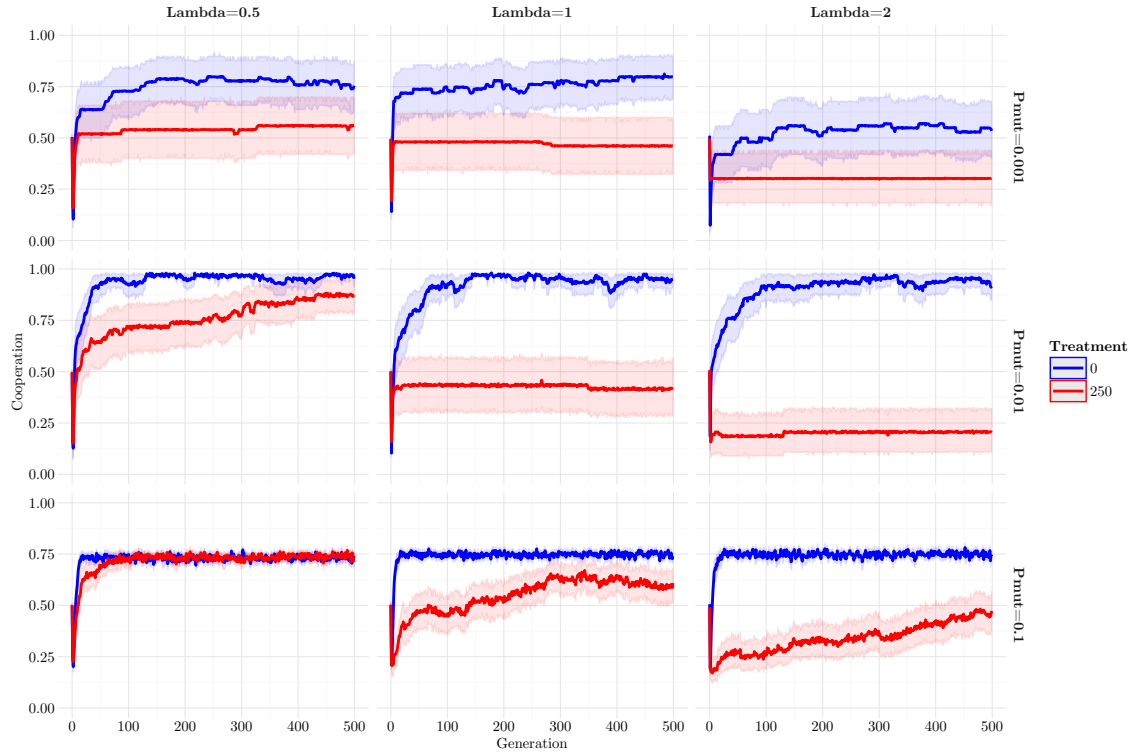
B.3   Simulations with Crossover

Figure B.3: Evolution of Cooperation within the Population. *Notes*: Fifty simulations of population of $n = 50$ chromosomes are conducted ($p_{mut} = 0.01$ and $\lambda = 1$). Each generation, 4% of the population is selected for crossover (hence $k = 2$). 95% bootstrapped confidence intervals around the average cooperation are superimposed.

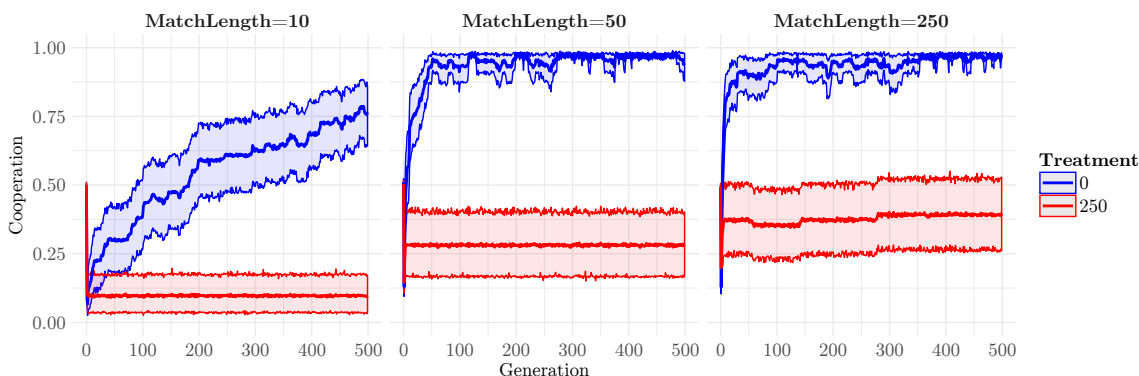## B.4   Simulations for Different Expected Match Lengths

Figure B.4: Evolution of Cooperation For Different Match Lengths. *Notes*: Fifty simulations of population of $n = 50$ chromosomes are conducted ($p_{mut} = 0.01$, $\lambda = 1$, and $k = 0$). 95% bootstrapped confidence intervals around the average cooperation are superimposed.

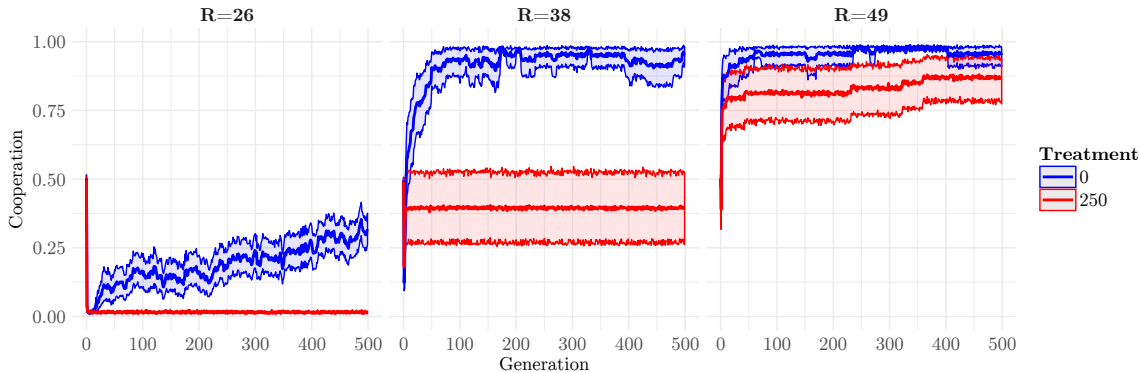## B.5   Simulations for Different Payoff Matrix

Figure B.5: Evolution of Cooperation For Different Values of R. *Notes*: R denotes benefit to cooperation. Fifty simulations of population of $n = 50$ chromosomes are conducted ($p_{mut} = 0.01$, $\lambda = 1$, and $k = 0$). 95% bootstrapped confidence intervals around the average cooperation are superimposed.

C Original Set of Experiments

Six sessions of the experiment were conducted, with sessions in May and September 2015. The following list summarizes the differences from the experiment described in the main body of the paper:

1. In the original set of experiments, each period lasted exactly four seconds.

2. In order to provide sufficient time to modify the strategies, we did not restrict the strategy construction to be done within one period. Instead, subjects incurred a cost of one point per second when rules were unlocked. Fixed costs for clicking the "Unlock Rules" button were the same as in the current experiment.

3. There was no split into two groups based on quiz scores.

4. The first five supergames were implemented using the game method. In other words, to get familiar with the game, subjects played the first five supergames via direct-response.

In the experiments presented in the main body of the paper, we chose to modify these four items in order to 1) provide a sufficient amount of time; 2) simplify the cost structure; 3) ensure that subjects who do not understand the interface do not influence the outcome; and 4) ensure that learning via direct-response under different costs had no effect on the initial propensity to cooperate via strategy-method. Table C.1 summarizes parameters and sessions for the original set of experiments.

Treatment	Stage Game Payoff	Continuation Probability	Sessions	N. Sub.	Description
$C = 0$		$\delta = .98$	2	34	*Costless* strategy adjustment within supergames
$C = 25$		C D C 38,38 12,50 D 50,12 25,25	$\delta = .98$	2	36	*Medium Costs* to strategy adjustment within suupergames
$C = 250$		$\delta = .98$	2	32	*High Costs* to strategy adjustment within suupergames

Table C.1: Treatments summary
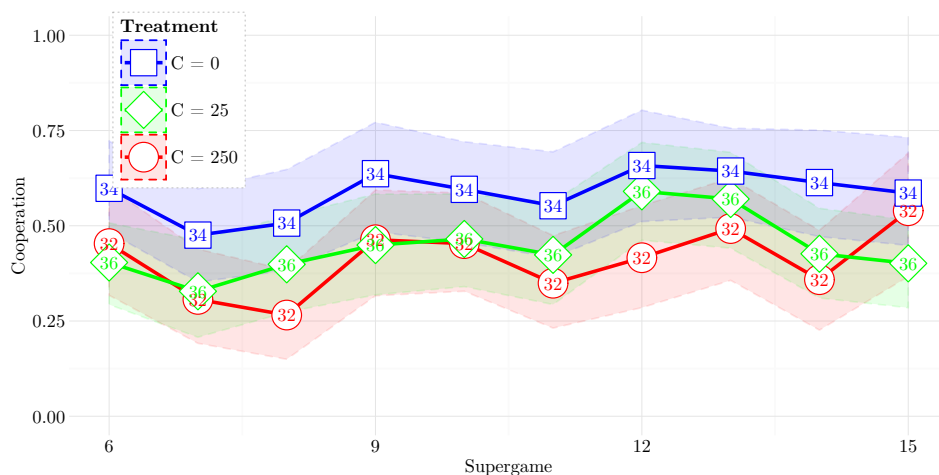
C.1 Cooperation

Figure C.1: Cooperation. *Notes*: Unit of observation is subject's average cooperation over all periods in the supergame. Number inside circle/rectangle represents number of subjects. 95% bootstrapped confidence intervals are superimposed.

Cooperation rates in the original set of experiments are comparable to those of the main experiment. There are some differences in the dynamics of cooperation for the $C = 0$ treatment. These differences may be attributed to including all subjects and having supergames 1-5 use the game method.

C.2 Strategy Estimates

Supergames	Cost	Tit For Tat	ALLD	Grim	Exp. Tit For Tat	2 Tits for 1 Tat	False Cooperator	ALLC	Lenient Grim 2	Lenient Grim 3	Tit For 2 Tats	Alternator	2 Tits for 2 Tats	Exp. Tit For 2 Tats	Exp. Tit For 3 Tats	Tit For 3 Tats	WSLS	WSLS 2 Puns.	Exp. Grim 3	β	Cooperation	# of Subjects
1-3	0	0.31*** (0.09)	0.09** (0.05)	0.04 (0.04)		0.08* (0.06)	0.17** (0.08)	0.04 (0.04)	0.04 (0.04)	0.10* (0.06)	0.04 (0.04)	0.02 (0.02)		0.03 (0.04)			0.04 (0.03)			0.88	0.50	34
	25	0.09* (0.06)	0.18*** (0.07)		0.20*** (0.08)		0.11* (0.07)	0.09** (0.05)	0.10* (0.07)					0.08** (0.05)	0.07 (0.05)				0.03 (0.03)	0.87	0.35	36
	250	0.19** (0.08)	0.28*** (0.09)	0.12** (0.07)	0.03 (0.03)	0.04 (0.05)	0.10* (0.06)	0.06* (0.04)			0.03 (0.04)	0.03 (0.03)				0.06 (0.05)	0.06* (0.05)			0.91	0.32	32
5-7	0	0.42*** (0.10)	0.18*** (0.07)		0.08* (0.05)		0.02 (0.03)		0.17** (0.09)	0.04 (0.04)		0.02 (0.03)		0.06* (0.04)						0.94	0.58	34
	25	0.13** (0.07)	0.23*** (0.07)	0.08 (0.09)	0.04 (0.04)	0.16 (0.12)		0.03 (0.03)				0.03 (0.02)	0.09* (0.06)			0.04 (0.05)	0.10* (0.07)			0.93	0.46	36
	250	0.18** (0.09)	0.19*** (0.07)	0.20** (0.09)	0.06 (0.05)		0.07* (0.05)	0.12* (0.08)		0.06 (0.06)	0.03 (0.04)	0.04 (0.04)	0.05 (0.05)							0.92	0.37	32
8-10	0	0.32*** (0.11)	0.17*** (0.07)		0.05 (0.04)	0.15* (0.10)	0.03 (0.03)		0.06 (0.05)	0.03 (0.05)	0.13** (0.08)			0.02 (0.03)			0.04 (0.03)			0.94	0.60	34
	25	0.20*** (0.07)	0.18*** (0.06)	0.11** (0.06)	0.12** (0.05)	0.04 (0.04)	0.01 (0.03)	0.08* (0.05)			0.05 (0.05)	0.03 (0.02)		0.06 (0.05)			0.03 (0.03)	0.03 (0.03)	0.05 (0.05)	0.89	0.42	36
	250	0.16** (0.08)	0.17** (0.07)	0.10 (0.10)	0.06* (0.04)	0.12 (0.10)	0.02 (0.03)	0.04 (0.05)	0.05 (0.05)	0.13* (0.09)	0.06 (0.06)		0.05 (0.06)				0.04 (0.03)			0.93	0.50	32

Table C.2: MLE Strategy Estimates. *Notes*: Bootstrapped standard errors are in parentheses. *, **, ***, denote significance at the 0.10, 0.05, and 0.01 levels, respectively. Labels 1-3, 5-7, etc. refer to strategy-method ordering (6-8, 10-12, etc. overall)