Mixed Strategies in the Indefinitely Repeated Prisoner's Dilemma

January 11th, 2019

Julian Romero[†] Yaroslav Rosokha[‡]

Abstract

We use an experimental approach to elicit mixed strategies from human subjects playing indefinitely repeated prisoner's dilemma. Our approach allows us to address the recent debate regarding whether subjects play mixed-strategies by providing direct empirical evidence. We find that although majority of subjects do use mixed strategies, the behavior is best summarized by three pure strategies: tit-for-tat, grim trigger, and always defect. We also find that a notable proportion of subjects use a class of mixed strategies that we refer to as the mixed-tit-for-tat class of strategies.

Keywords: Indefinitely Repeated Prisoner's Dilemma, Mixed Strategies, Experimental Design, Strategy Elicitation

† Eller College of Management, Univeristy of Arizona • jnromero@email.arizona.edu
‡ Krannert School of Management, Purdue University • yrosokha@purdue.edu

1 Introduction

The repeated prisoner's dilemma has been used to study a variety of topics in economics.[1] Theoretical work has largely focused on folk theorems, which show that as long as players are sufficiently patient, then any payoff above a minimum threshold can be obtained as an equilibrium payoff. Experimental work has provided an important complement to the theory by testing which of the plethora of equilibria will emerge in a given setting. Analysis of directly observable outcomes in experiments (such as cooperation and defection) is relatively straightforward, while analysis of the underlying (non-observable) strategies that generate these outcomes is more difficult and often requires assumptions about the set of strategies that subjects use. Because of this, some studies have found strong evidence that subjects use mixed strategies while others have found strong evidence that subjects use pure strategies. This paper aims to directly test whether subjects use mixed repeated-game strategies in the indefinitely repeated prisoner's dilemma.

In recent years, great strides have been made to better understand the strategies that subjects play in the repeated prisoner's dilemma (see Section 2.5 of Dal Bó and Fréchette, 2018, for a review). This work has typically either estimated strategies from observed actions (e.g., Dal Bó and Fréchette, 2011) or elicited strategies directly (e.g., Dal Bó and Fréchette, 2017). The previous studies, however, have largely focused on pure strategies.[2] Recent work by Breitmoser (2015) suggests that a large majority of subject behavior in the indefinitely repeated prisoner's dilemma games can be explained by a single mixed strategy called Semi-Grim (henceforth SG). In this paper, we use a novel experimental approach to directly elicit mixed repeated-game strategies. This allows us to empirically validate whether subjects use mixed strategies in the indefinitely repeated prisoner's dilemma. We find that while a majority of subjects do use mixed strategies, the majority of behavior is explained with three pure strategies: Tit-For-Tat, Grim Trigger, and Always Defect (henceforth TFT, GRIM, and ALLD, respectively).

Our experimental design is the first (to our knowledge) that allows for elicitation of history contingent mixed-strategies in repeated games. Previous studies that have elicited repeated game strategies (Dal Bó and Fréchette, 2017; Embrey, Mengel, and Peeters, 2016; Romero and Rosokha, 2018) have asked subjects to specify an action to be played after a given history. In our experiment, a subject specifies a probability (up to 2 decimal places) with which to play one of the two actions (C or D) contingent on the actions in the previous period. For example, a subject could specify to play C with probability 70% if DC was played in the previous period. The design allows subjects to construct mixed repeated-game strategies in which they have different probabilities of playing C or D in the first period and after each of the four possible memory one histories. Therefore, there are 101^5 possible strategies available to subjects. We use this approach to conduct an experiment on the indefinitely repeated prisoner's dilemma with a continuation probability of $\delta = .95$.

[1]Examples include: quantity setting oligopolies (Mailath and Samuelson, 2006), R&D races (Cockburn and Henderson, 1994), trade wars (Maggi, 1999), international relations (Powell, 1993), and labor negotiations (Kahn, 1993).

[2]One exception is Dal Bó and Fréchette (2017) who have a treatment in which subjects can choose from a predefined set of strategies. One option in the set is a strategy that randomizes with the same (subject specified) probability in every period regardless of the history.

We have several main findings from these experiments. First, we find no difference in cooperation levels between the current experiment and previous strategy elicitation experiments that have only elicited pure-strategies. This suggests that the mixed-strategy elicitation does not have a significant impact on subject behavior. Second, we find that a majority of subjects (68 percent) use mixed strategies. Third, even though most subjects do use mixed strategies, commonly studied pure strategies (TFT, GRIM, ALLD) still capture a majority (59 percent) of behavior. Finally, in addition to these commonly studied pure strategies, we find that a notable proportion (30 percent) of behavior is captured by a class of mixed-strategies that we refer to as the mixed-tit-for-tat class of strategies. Strategies in this class cooperate after CC, defect after DD, mix after CD and DC, and cooperate with higher probability after DC than CD. In this class there are two types of strategies; those which cooperate in the first period (which we refer to as mixed-TFT strategies) and those that defect in the first period (which we refer to as mixed-STFT strategies). These types of strategies are similar to the previously studied mixed-strategy SG in that they play C after CC, D after DD, and randomize after CD and DC. However, these types of strategies differ from SG in that they have different probabilities of cooperation after CD and DC (in the direction of TFT) and they specify behavior in the first period.

Our work contributes to three streams of literature. The first, and closest stream of literature experimentally studies strategy choice in the indefinitely repeated prisoner's dilemma. The papers in this literature either estimate strategies using indirect inference methods (Dal Bó and Fréchette, 2011; Fudenberg, Rand, and Dreber, 2012; Camera, Casari, and Bigoni, 2012; Bigoni, Casari, Skrzypacz, and Spagnolo, 2015); elicit strategies by requiring subjects to "construct" the strategies that they want to use (Dal Bó and Fréchette, 2017; Romero and Rosokha, 2018); or choose from a predefined set of strategies to be played in supergames (Dal Bó and Fréchette, 2017; Cason and Mui, 2017). While each of the approaches has it's benefits and drawbacks, the strategies in the above papers have been restricted to pure strategies. In this paper, we allow subjects to construct mixed strategies, which enables us to empirically assess whether focusing on pure strategies is restrictive in the context of the indefinitely repeated prisoner's dilemma.

The second stream of literature studies mixed strategies in the repeated prisoner's dilemma. Ely and Välimäki (2002) and Ely, Hörner, and Olszewski (2005) theoretically characterize a set of belief-free equilibria that yield payoffs consistent with the Folk theorems. More recently, Breitmoser (2015) theoretically characterized a particular class of strategies which he termed Semi-Grim. In addition, Breitmoser (2015) uses the data from 17 treatments from four experiments to estimate these strategies and finds that semi-grim captures the behavior of majority of the subject (at least 50% in 14 out of 17 treatments).[3] In this paper, we contribute to this literature by providing direct empirical evidence on whether people actually use mixed-strategies, and if so, what type of mixed-strategies do they use.

The third stream of literature to which we contribute is a fairly small set of studies that

[3]Backhaus and Breitmoser (2018) does a further analysis of data from a wide range of repeated prisoner's dilemma experiments (32 treatments) using data-mining techniques. They still find strong support that a majority of subjects behavior can be explained by SG.

experimentally investigate mixed strategies in broader repeated games. In particular, Bloomfield (1994); Ochs (1995); Shachat (2002); Chmura and Güth (2011); Noussair and Willinger (2011); Cason, Friedman, and Hopkins (2013) elicit mixed-strategies in repeated game setting by providing the subjects with the randomization device to specify a mixed-strategy. Our work is different in several respects. First, the above papers study mixed strategies in a setting with a unique mixed-strategy equilibrium (the exception is Chmura and Güth (2011), who study the setting with a unique symmetric mixed-strategy equilibrium but multiple asymmetric equilibiria in pure strategies). Second, we allow subjects to specify mixed-strategies that condition on the prior history of play, which leads us to investigate more complex repeated-game strategies. Lastly, we focus on the indefinitely repeated prisoner's dilemma.

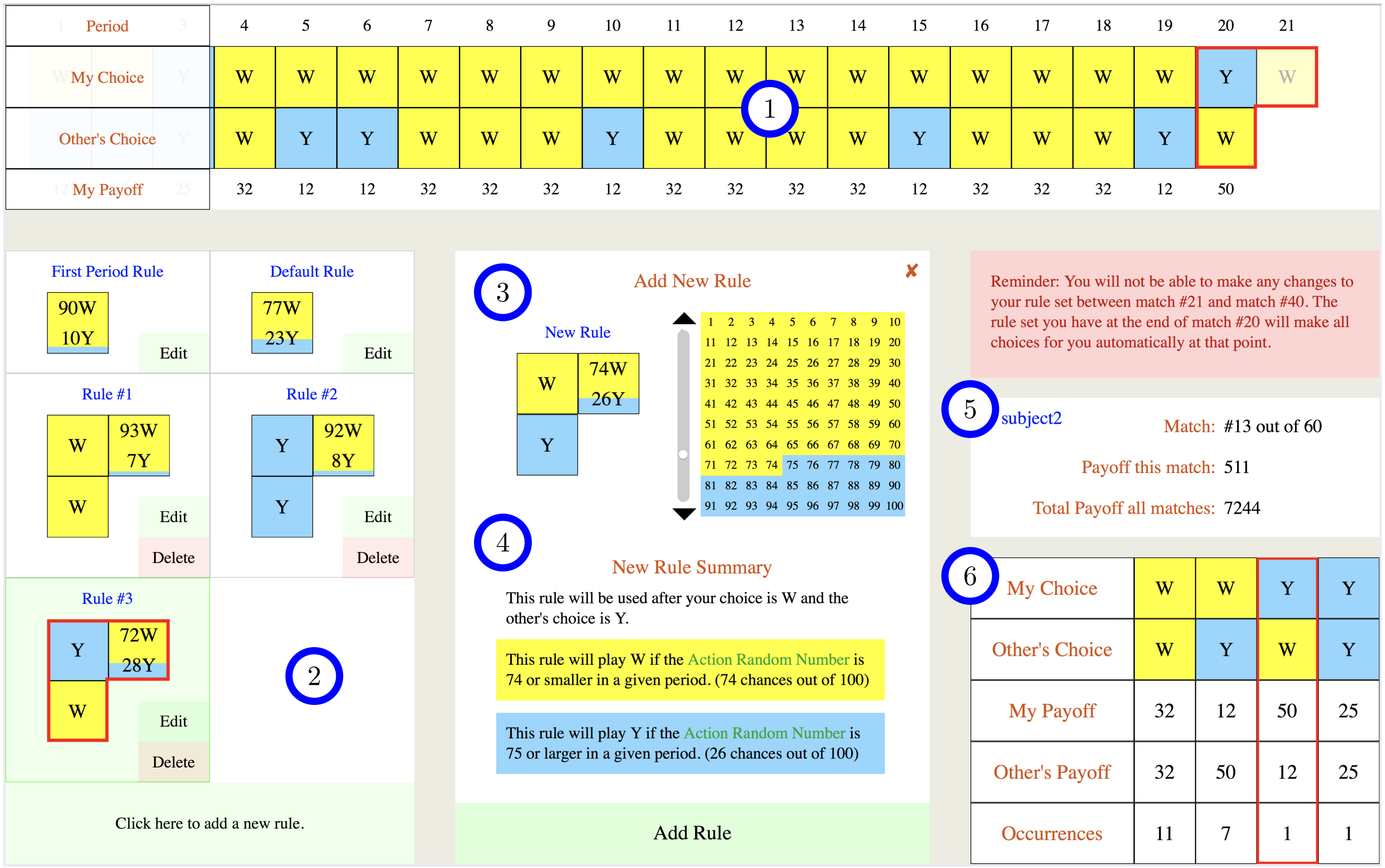
The rest of the paper is organized as follows: Section 2 presents details of our experimental design. Section 3 presents the results. Lastly, Section 4 provides a concluding discussion of the results and outlines directions for future research.

2 Experimental Design

In this section, we describe the interface and the experimental design. Specifically, we modify the experimental interface of Romero and Rosokha (2018) to allow elicitation of memory-1 mixed strategies. Figure 1 presents a screenshot of the experimental interface.[4] Next, we highlight the important aspects of the experimental interface and design.

[4]The interface was developed using a Simple Toolbox for Experimental Economics Programs (STEEP). Further information about the interface can be found at `http://jnromero.com/research/mixedStrategyChoice`.

Figure 1: Screenshot of the Experimental Interface.



Notes: The neutral action names W and Y correspond to the usual action names C and D from the prisoner's dilemma. The screenshot shows: (1) History, (2) Rule Set, (3) Rule Constructor, (4) New Rule Summary, (5) General Information, (6) Payoff Table.

2.1 Rules and Strategies

The main component of our experimental interface is the ability to construct strategies using a set of "if [input]-then [output]" rules. The input of a rule is an action profile in at most one previous period, and the output of a rule is the probability with which to play a particular action. Subjects are able to modify strategies by adding and subtracting rules from their rule set. The rule set will then make a choice for a subject in a given period based on the history. The choice is determined by the rule that has the same input as the last period of the history. If a rule set does not contain a rule that has the last period of the history as an input, then the default rule will be used to make the choice. Subjects are required to specify both a default rule and a first period rule before their rule set makes any choices for them, which ensures that the rule set is able to specify an action after every history.

There are two main differences between the current experiment and Romero and Rosokha

(2018). First and foremost, in the current experiment, we allow subjects to specify rules with probabilistic outputs. This modification allows us to study mixed strategies in the indefinitely repeated prisoner's dilemma, which is the main goal of this paper. Second, we restrict subjects to memory-1 rules (no more than 1 period as an input). This modification makes the strategy elicitation process less complex while still allowing subjects to construct strategies of interest from Dal Bó and Fréchette (2018) (such as GRIM, TFT, STFT, ALLD, and ALLC), SG from Breitmoser (2015), and belief-free equilibrium strategies from Ely, Hörner, and Olszewski (2005)).

Our experimental interface allows subjects to construct rules with probabilistic outputs by using the rule constructor (#3 in Figure 1). Specifically, subjects can use a slider to specify a cutoff between 0 and 100.[5] The cutoff determines the probability with which C is selected. The way we implement randomization, and explain it to subjects, is by drawing an "action random number" each period. The action random number is an integer between 0 and 100 (inclusive). If the integer is less than or equal to the cutoff, then C is played. If the integer is greater than the cut-off, then D is played. Subjects are reminded that each one of them receives their own independent draw of the action random number in each period.

Figure 1 (# 2) presents an example of a rule set that can be constructed with the interface.[6] We will denote this rule set as $\{FP \rightarrow 90; \rightarrow 23; CC \rightarrow 93; DD \rightarrow 8; DC \rightarrow 72\}$. Given this strategy, the subject will cooperate with probability 90 percent in the first period, and will cooperate with probability 93, 23, 72, 8 percent if CC, CD, DC, DD was played in the previous period, respectively. Note that since there is no rule with input CD, then the default rule will be used to make the choice after that action profile.

2.2 Experimental Protocol

The experiment consisted of three sessions run at the Vernon Smith Experimental Economics Laboratory at Purdue University in April 2018. Each session consisted of instructions, incentivized quiz to ensure that subjects understood the instructions, and 60 supergames. All payoffs were displayed in Experimental Currency Units (ECUs) and were converted at the end of the experiment at the exchange rate of 1500 ECUs to 1.00 USD. Next, we describe specific parts of the experimental design in more detail.

[5] When a subject decides to create a rule, the rule constructor appears with "?"s in each box. When the subject clicks on one of the boxes corresponding to the rule input, the box changes from the question mark to either "W" or "Y" randomly. If the box already has a "W" or a "Y" then it changes to the other action when clicked. To set the output of the rule, the subject needs to click the slider next to the rule. There is no default value for the slider. The marker on the slider is not visible until the subject clicks on it. Once the subject has clicked the slider, the proportion of squares corresponding to the probability of playing "W" are colored yellow, the proportion of squares corresponding to the probability of playing "Y" are colored blue, and the corresponding numbers are summarized in the rule output square. In addition, a written summary of the rule is displayed below the constructed rule. Detailed screenshots of this process are presented in Figure A-1 in the Appendix.

[6] Figure A-2 in the Appendix provides a few more examples of strategies that can be constructed with our interface, including TFT, GRIM, SG, and mixed-TFT.

2.2.1 Game Parameters

We picked the parameters for the experiment to match those of Romero and Rosokha (2018) and one treatment of Dal Bó and Fréchette (2017). Specifically, we used the stage game payoffs that are displayed in Figure 1 (# 6) and the continuation probability $\delta = 0.95$.[7] These parameters allow a direct comparison to Romero and Rosokha (2018) and one treatment of Dal Bó and Fréchette (2017).

2.2.2 Instructions and Quiz

Instructions used in this experiment consisted of a sequence of interactive screens which explained all of the aspects of the experiment and details of the experimental interface. Throughout the instructions there were 20 quiz questions. The quiz was incentivized as follows. Subjects earned \$5.00 if they answered at least 15 out of 20 questions correctly and \$0.00 otherwise. Among the 76 subjects who participated in the experiment, 56 passed the quiz.[8] The subjects who passed the quiz were randomly matched into groups of 8-12 subjects.[9]

2.2.3 Experimental Stages

Similar to Dal Bó and Fréchette (2017) and Romero and Rosokha (2018), there were three types of supergames: Direct-response, Non-binding, and Locked-Response. Next, we briefly describe each of the stages and the purpose it serves.

Direct-Response Stage (Supergames 1-10). In the direct-response stage subjects play the game by choosing C or D each period. The direct response stages ensures that subjects learn about the strategic tension in the game, without having to specify strategies.[10]

[7] We used two sequences of 60 supergame lengths that were pre-drawn using the computer according to continuation probability $\delta = 0.95$. In each session, there were two groups, with one group receiving the first realization and the other receiving the second realization. The realizations are provided in Table A-1 in the Appendix.

[8] Comparing demographic variables for those that passed the quiz and those that did not, the statistic that stood out between was whether the student attended high-school outside of the US. In particular 11 (out of 20) of those that did not pass the quiz attended high-school outside of the US, indicating that one of the reasons for doing poorly on the quiz may have been understanding of the language.

[9] To keep group sizes relatively close in size, we decided to have a minimum group size of 8 and a maximum group size of 14. Therefore, if only 12 subjects passed the quiz, they were all matched in the same group, but if 16 subjects passed the quiz, then they were divided into 2 groups of 8. In our experiment, each session ended up with exactly two groups, the smallest of which contained 8 subjects and the largest of which contained 12 subjects.

[10] One difference between Romero and Rosokha (2018) and the current experiment is that we require subjects to confirm their opponents action after each period. More specifically, they received the following message: "To continue click the key corresponding to the choice of the participant that you are matched with from the previous period on the keyboard (either W or Y)." We did this to avoid situations like the following: suppose two subjects were playing D for many periods, and then one subject plays C for one period, and the other subject quickly continues to play D without necessarily processing that their opponent had played C in the previous period. This confirmation was added to ensure that subjects correctly processed the choice of their opponent before making their choice in the next period. This design allows the subject to progress quickly if their opponent plays as expected, but must pause to process the choice if their opponent plays contrary to what was expected.

Non-Binding Stage (Supergames 11-20). During the non-binding stage, subjects were provided up to 10 minutes to construct the initial set of rules and up to 2 minutes before each additional supergame.[11] Importantly, there was no time limit on the duration of each period during a supergame. As the supergame progressed in the non-binding stage, subjects were informed of the action that their rule set would play each period given their draw of the Action Random Number. Subjects were given the option to manually deviate from the prescription of their rule set in every period of the non-binding stage. When subjects deviated from the prescription of their rule set, they were given a warning that reminded them that in the locked-response stage, their rule set would automatically make their choices for them.

Locked-Response Stages (Supergames 21-40 and Supergames 41-60). In the locked response stage, subjects' rule sets made choices for them automatically. Subjects were not able to change their rule sets during the locked-response stage. The current experiment consisted of two locked-response stages (as opposed to only one in Romero and Rosokha (2018)), and subjects were given up to 10 minutes to edit their rule sets between the locked-response stages. The locked-response stage served as an incentive to construct (and understand) strategies during the non-binding stage. We decided to include the second locked-response stage to ensure that subjects had sufficient time and experience to evaluate mixed strategies. In addition, comparing strategies between the first and second locked-response stage allows us to assess the evolution of strategies.

3 Experimental Results

The results section consists of one remark regarding the experimental design and four main results regarding the strategies that subjects play. First, since the elicitation of mixed strategies is relatively complex, we want to ensure that subjects' behavior is in line with prior studies that used similar parameters. To address this, in Section 3.1, we provide a direct comparison of cooperation rates between our experiments and Romero and Rosokha (2018) (which elicits only pure strategies) and find no significant difference between the two studies. Second, the main goal of this experiment is to examine whether subjects play mixed strategies in the indefinitely repeated prisoner's dilemma. To address this, in Section 3.2, we analyze the elicited strategies and find that while most subjects use mixed strategies, the majority of behavior is explained by three commonly studied pure strategies: TFT, GRIM, and ALLD. In addition, we find that a notable proportion of subjects used strategies from the mixed-tit-for-tat class of strategies. Lastly, we analyze the evolution of strategies in our experiment. Specifically, in Section 3.3 we investigate changes that subjects made between the two locked-response stages and find evidence that the strategies are becoming less random.

[11] These time limits were never close to binding. See Figure A-3 in the Appendix for details.

3.1 Cooperation

Table 1 presents the average cooperation rate observed for mixed-strategy elicitation stages in the current experiment (labeled *Current*) and pure strategy elicitation stages in Romero and Rosokha (2018) (labeled *RR2018*). The cooperation rates are divided into groups of supergames based on the experimental design. In particular, supergames 11–20 were non-binding, and supergames 21–40 were locked response stages in both experiments.[12] For each of these groups of supergames, we present the cooperation rates for the first period, first 4 periods, last 4 periods, and all periods.The cooperation rate is computed by first averaging cooperation in each of the corresponding periods in each supergame, and then averaging across the corresponding supergames.

Table 1: Average Cooperation Rate.

Experiment	Current		RR2018	Current		RR2018	Current
Subjects	56		44	56		44	56
Supergames	11–20		11–20	21–40		21–40	41–60
Type	NB		NB	LR		LR	LR
First Periods	0.67 (0.05)	∼	0.75 (0.06)	0.7 (0.05)	∼	0.77 (0.06)	0.72 (0.05)
First 4 Periods	0.65 (0.04)	∼	0.67 (0.04)	0.61 (0.04)	∼	0.66 (0.05)	0.59 (0.04)
Last 4 Periods	0.62 (0.03)	∼	0.58 (0.04)	0.49 (0.04)	∼	0.54 (0.04)	0.49 (0.04)
All Periods	0.63 (0.03)	∼	0.61 (0.04)	0.53 (0.04)	∼	0.58 (0.05)	0.53 (0.04)

Notes: The unit of observation is the average cooperation rate by a subject over the corresponding range of supergames. The cooperation rate is the fraction of periods that a subject cooperated in the given range of periods. Bootstrapped standard errors are in parentheses. If the supergame is less than four periods, then the cooperation rate for the first four and the last four is set to the cooperation rate for all periods. Supergames 11–20 correspond to the non-binding stage (labeled NB); supergames 21–40 and 41–60 correspond to the locked-response stage (labeled LR). Tests between treatments are carried out using non-parametric permutation tests (Good, 2013).

Table 1 shows that cooperation rates from the current experiment are remarkably close to the cooperation rates from the prior study when looking at the cooperation rates across different periods. In addition, Figure A-4 in the Appendix compares the current experiment to a treatment

[12] Note that Romero and Rosokha (2018) had two treatments. Treatment 1 had direct-response for supergames 1–10, non-binding for supergames 11–20, locked-response for supergames 21–50, and direct response for supergames 51–60. Treatment 2 had direct-response for supergames 1–20, non-binding for supergames 21–30, locked response for supergames 31–60. The current experiment had direct-response for supergames 1–10, non-binding for supergames 11–20, locked response for supergames 21–40, and then another locked response for supergames 41–60. If we focus on strategy elicitation stages, we can directly compare supergames 11–20 and supergames 21–40 of the current experiment to Treatment 1 of Romero and Rosokha (2018) (44 subjects). Finally, there is no valid comparison for supergames 41–60 as both treatments in Romero and Rosokha (2018) had a different progression of stages and opportunities for strategy revision up to supergame 41.

from Dal Bó and Fréchette (2017) that uses the same parameters (yet a different interface and design) and provides further evidence that cooperation levels seen in this experiment are in line with previous studies that elicit pure strategies. Remark 1 summarizes these results.

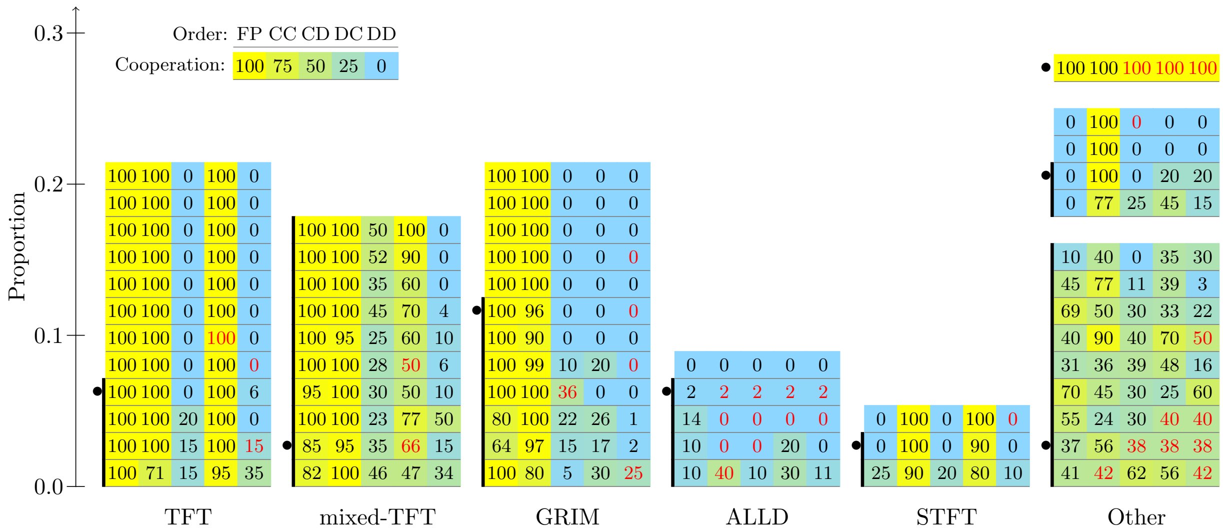
Remark 1 *Average cooperation rates are consistent with experiments that elicit pure strategies.*

3.2 Strategies

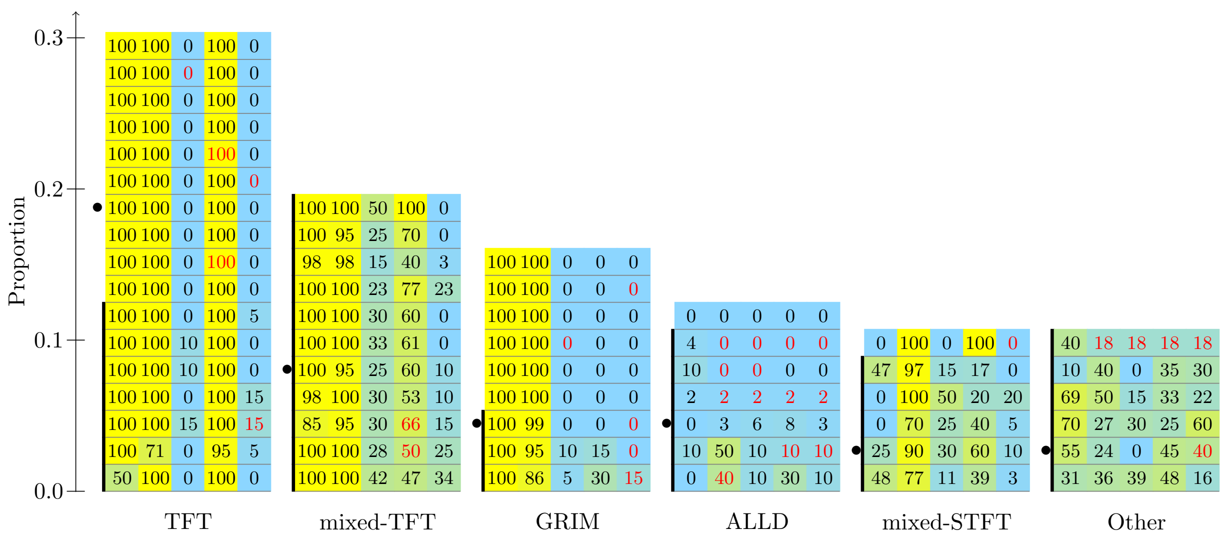
The main focus of this paper which is analyzing the elicited strategies. Figure 2 presents all of the strategies observed in the two locked-response stages of the current experiment. The strategies are organized using a clustering approach which groups similar strategies together without specifying any categories beforehand.[13] The analysis of the strategies leads to three main results.

[13]We use affinity propagation clustering approach (Frey and Dueck, 2007) with Euclidean distances between strategy vectors as the similarity criteria. A strategy vector consists of the five cooperation percentages corresponding to the histories ($\emptyset, CC, CD, DC, DD$).

Figure 2: Elicited Strategies



(a) First Locked-response Stage



(b) Second Locked-response Stage

Notes: **(a)** Strategies elicited during the first locked-response stage. **(b)** Strategies elicited during the second locked-response stage. Strategies are presented as vectors of cooperation percentages after each of the five histories ($\emptyset, CC, CD, DC, DD$). Red numbers denote a situation in which proportion of cooperation is determined by the default rule. A black line on the left hand side of a strategy denotes that the strategy is a mixed strategy. A black bullet on the left hand side of a strategy denotes that the strategy is the cluster exemplar. Clusters with an exemplar that has average RRM less than 5.0 are labeled with pure strategies.

The first result is regarding the use of mixed strategies:

Result 1 *A majority of subjects use mixed strategies.*

Figure 2 shows that 38 (out of 56) subjects specify a mixed strategy. Specifically, the figure highlights all of the mixed-strategies with a black bar on the left hand side. To further examine the randomness in these strategies, we define the *rule randomness measure* (RRM) to be two times the deviation from the closest pure strategy output. For example, a rule with output 80 would have an RRM of 40 ($= 2 * (100 - 80)$), a rule with output of 10 would have an RRM of 20 ($= 2 * (10 - 0)$. Furthermore, any rule with pure strategy output would have an RRM of 0, a rule with completely mixed output would have an RRM of 100, and an output drawn from a uniform distribution would have an expected RRM of 50. We find that the average RRM in the first period (14.4), after CC (15.1), and after DD (14.4) are all lower compared to those after CD (22.9) and DC (30.4). These results suggest that behavior in the first period and after CC and DD histories are less random, than the behavior after the DC and CD histories.[14] It is noteworthy that measures across all five strategy components (FP, CC, CD, DC, DD) are relatively low compared to random behavior, suggesting that specified mixed strategies are not too far removed from pure strategies.

To better identify regularities in the strategies, we run clustering analysis to group similar strategies together. Figure 2 shows clusters of strategies identified in our experiment. Each cluster is characterized by an exemplar – the most representative member among all the strategies in that cluster. In the figure, we mark each exemplar with a black bullet on the left hand side of the strategy. Based on the cluster exemplar, we identify five major clusters.

The second result is regarding the types of strategies that subjects used as they relate to commonly studied pure strategies:

Result 2 *A majority of behavior is captured by three pure strategies: TFT, GRIM, ALLD.*

We find three clusters that have exemplars with an average RRM of less than 5, and combined contain 59% of subjects. The first cluster has 17 (out of 56) subjects, has an exemplar of $\{FP \to 100; CC \to 100; CD \to 0; DC \to 100; DD \to 0\}$, has an average RRM of 0, and is therefore labeled TFT. The second cluster has 9 (out of 56) subjects, has an exemplar of $\{FP \to 100; CC \to 99; CD \to 0; DC \to 0; DD \to 0\}$, has an average RRM of 0.4, and is therefore labeled GRIM. Finally, the third cluster has 7 (out of 56) subjects, has an exemplar of $\{FP \to 0; CC \to 3; CD \to 6; DC \to 8; DD \to 3\}$, has an average RRM of 4, and is therefore labeled ALLD. These three cluster represent pure strategies that have been consistently found in literature. For example, Dal Bó and Fréchette (2018) find that these three strategies account for the majority of behavior in 15 out of 17 experimental treatments across a variety of parameters.

The third result is regarding the types of mixed-strategies that subjects used:

Result 3 *A notable proportion of behavior is captured by the mixed-tit-for-tat class of strategies.*

[14]Using a Wilcoxon signed-rank test we obtain p-values < 0.01 for all combinations in $\{FP, CC, DD\} \times \{DC\}$; p-values < 0.05 all combinations in $\{CC, DD\} \times \{CD\}$; and p-value of 0.06 for the (FP, CD) combination.

We find two clusters which have strategies that belong in the mixed-tit-for-tat class of strategies. The first cluster has 11 (out of 56) subjects, has an exemplar of $\{FP \rightarrow 100; CC \rightarrow 95; CD \rightarrow 25; DC \rightarrow 60; DD \rightarrow 10\}$, has an average RRM of 32. This cluster is labeled mixed-TFT. The second cluster has 6 (out of 56) subjects, has an exemplar of $\{FP \rightarrow 25; CC \rightarrow 90; CD \rightarrow 30; DC \rightarrow 60; DD \rightarrow 10\}$, has an average RRM of 46. This cluster is labeled mixed-STFT. The strategies in these two clusters all cooperate with high probability after CC, all defect with high probability after DD, and almost all (16 out of 17) cooperate with higher probability after DC than CD. The strategies in the two clusters differ in their probabilities of cooperation in the first period. The larger mixed-TFT cluster has cleaner trends compared to the smaller mixed-STFT cluster. More specifically, strategies in the mixed-TFT cluster have relatively low RRM for FP, CC and DD (averages of 3.5, 3.1, and 21.8, respectively) and are consistent with the TFT strategy. In addition, all strategies in this cluster have a higher probability of cooperating after DC then CD (median difference of 30 percent). Lastly, the probability of cooperating after DC is greater than or equal to 50 in all but two strategies in this cluster (those that have 40 and 47) and the probability of cooperating after CD is always less than or equal to 50. Similar trends hold in the smaller mixed-STFT cluster, although the behavior more is random. Interestingly, strategies in the mixed-tit-for-tat class of strategies cooperate with positive probability after CD, which is in contrast to commonly studied memory-1 strategies such as ALLD, TFT, GRIM, STFT, and WSLS which all defect after CD.

3.3 Evolution of Strategies

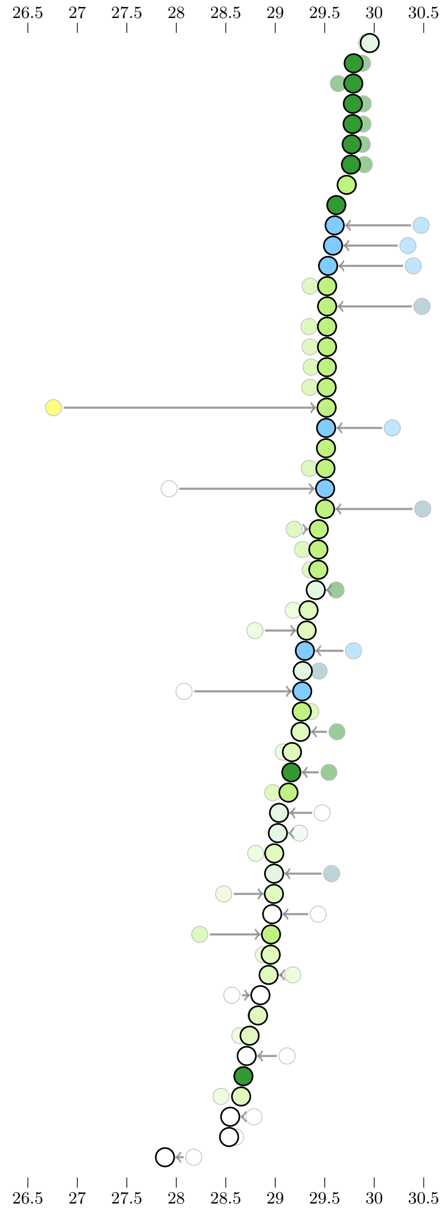
The two locked-response stages of our design allow us to assess the evolution of strategies in the experiment. Figure 3 presents data on all of the changes that subjects made to their strategies. Panel (a) presents strategies in the two locked response stages (labeled LR1 and LR2) as well as the average RRM for each strategy. We use arrows to identify if a change has been made and colors to identify the resulting change in the average RRM (red/black/green color denote an increase/no change/decrease in randomness). Panel (b) presents clusters that each strategy has been classified into in each of the two locked-response stages, with arrows denoting a transition to a different cluster.[15] Panel (c) shows the performance of a given strategy against the population of elicited strategies. The opaque circles denote the performance in the first locked-response stage and the solid circles denote the performance in the second locked-response stage.

[15]Note that we consider STFT and mixed-STFT as the same cluster.

Figure 3: Changes Between Two Locked-response Stages

LR1 FP	LR1 CC	LR1 CD	LR1 DC	LR1 DD	LR1 $\overline{RRM}$		LR2 FP	LR2 CC	LR2 CD	LR2 DC	LR2 DD	LR2 $\overline{RRM}$	Cluster LR1		Cluster LR2
0	100	0	100	0	0		0	100	0	100	0	0	STFT		mixed-STFT
100	100	0	0	0	0		100	100	0	0	0	0	GRIM		GRIM
100	100	36	0	0	14	⟶	100	100	0	0	0	0	GRIM		GRIM
100	100	0	0	0	0		100	100	0	0	0	0	GRIM		GRIM
100	100	0	0	0	0		100	100	0	0	0	0	GRIM		GRIM
100	100	0	0	0	0		100	100	0	0	0	0	GRIM		GRIM
100	100	0	0	0	0		100	100	0	0	0	0	GRIM		GRIM
0	100	0	90	0	4	⟶	50	100	0	100	0	20	STFT	⟶	TFT
100	96	0	0	0	1	⟶	100	99	0	0	0	0	GRIM		GRIM
0	0	0	0	0	0		0	0	0	0	0	0	ALLD		ALLD
14	0	0	0	0	5	⟶	4	0	0	0	0	1	ALLD		ALLD
2	2	2	2	2	4		2	2	2	2	2	4	ALLD		ALLD
100	100	0	100	0	0		100	100	0	100	0	0	TFT		TFT
0	100	0	0	0	0	⟶	100	100	0	100	0	0	DGRIM	⟶	TFT
100	100	0	100	0	0		100	100	0	100	0	0	TFT		TFT
100	100	0	100	0	0		100	100	0	100	0	0	TFT		TFT
100	100	0	100	0	0		100	100	0	100	0	0	TFT		TFT
100	100	0	100	0	0		100	100	0	100	0	0	TFT		TFT
100	100	100	100	100	0	⟶	100	100	0	100	0	0	ALLC	⟶	TFT
10	0	0	20	0	12	⟶	10	0	0	0	0	4	ALLD		ALLD
100	90	0	0	0	4	⟶	100	100	0	100	0	0	GRIM	⟶	TFT
100	100	0	100	0	0		100	100	0	100	0	0	TFT		TFT
41	42	62	57	42	82	⟶	0	3	6	8	3	8	Other	⟶	ALLD
0	100	0	0	0	0	⟶	100	100	0	100	0	0	DGRIM	⟶	TFT
100	100	20	100	0	8	⟶	100	100	10	100	0	4	TFT		TFT
100	100	0	100	6	2	⟶	100	100	0	100	5	2	TFT		TFT
100	100	0	100	0	0	⟶	100	100	10	100	0	4	TFT		TFT
64	97	15	17	2	29	⟶	47	97	15	17	0	32	GRIM	⟶	mixed-STFT
100	100	35	60	0	30	⟶	100	100	30	60	0	28	mixed-TFT		mixed-TFT
100	100	52	90	0	23	⟶	100	100	33	61	0	28	mixed-TFT		mixed-TFT
10	40	10	30	11	40	⟶	0	40	10	30	10	36	ALLD		ALLD
0	77	25	45	15	43	⟶	0	70	25	40	5	40	DGRIM	⟶	mixed-STFT
40	90	40	70	50	68	⟶	10	50	10	10	10	36	Other	⟶	ALLD
100	100	0	100	0	0	⟶	100	100	0	100	15	6	TFT		TFT
80	100	22	26	1	27	⟶	98	98	15	40	3	24	GRIM	⟶	mixed-TFT
95	100	30	50	10	38	⟶	98	100	30	53	10	35	mixed-TFT		mixed-TFT
100	99	10	20	0	12	⟶	100	95	10	15	0	12	GRIM		GRIM
100	100	15	100	15	12		100	100	15	100	15	12	TFT		TFT
45	77	11	39	3	48	⟶	48	77	11	39	3	49	Other	⟶	mixed-STFT
25	90	20	80	10	34	⟶	25	90	30	60	10	46	STFT		mixed-STFT
100	100	50	100	0	20		100	100	50	100	0	20	mixed-TFT		mixed-TFT
0	100	0	20	20	16	⟶	0	100	50	20	20	36	DGRIM	⟶	mixed-STFT
100	100	23	77	50	38	⟶	100	100	23	77	23	27	mixed-TFT		mixed-TFT
10	40	0	35	30	46		10	40	0	35	30	46	Other		Other
100	71	15	95	35	33	⟶	100	71	0	95	5	15	TFT		TFT
100	100	45	70	4	31	⟶	100	95	25	70	0	24	mixed-TFT		mixed-TFT
100	100	28	50	6	33	⟶	100	100	28	50	25	41	mixed-TFT		mixed-TFT
37	57	38	38	38	77	⟶	40	18	18	18	18	44	Other		Other
100	95	25	60	10	32		100	95	25	60	10	32	mixed-TFT		mixed-TFT
85	95	35	66	15	41	⟶	85	95	30	66	15	39	mixed-TFT		mixed-TFT
31	36	39	48	16	68		31	36	39	48	16	68	Other		Other
100	80	5	30	25	32	⟶	100	86	5	30	15	25	GRIM		GRIM
82	100	46	47	34	58	⟶	100	100	42	47	34	49	mixed-TFT		mixed-TFT
69	50	30	33	22	66	⟶	69	50	15	33	22	60	Other		Other
55	24	30	40	40	71	⟶	55	24	0	45	40	61	Other		Other
70	45	30	25	60	68	⟶	70	27	30	25	60	60	Other		Other

(a) Strategies **(b) Clusters**



(c) Performance

Notes: LR1 – first locked-response stage; LR2 – second locked-response stage. All panels are sorted by the performance of the strategy in LR2 against the population of other strategies in LR2. **(a)** Strategies are presented as vectors of cooperation percentages after each of the five histories ($\emptyset, CC, CD, DC, DD$). $\overline{RRM}$ is the average rule randomness measure. Arrows denote that subject modified her strategy between LR1 and LR2. Red, green, and black arrows denote an increase, decrease, and no change in $\overline{RRM}$, respectively. For LR2, an increase (decrease) in cooperation is highlighted in yellow (blue), with darker colors representing greater change. **(b)** Clusters identified in each of the locked-response stages. Arrows denote a transition into a different cluster. **(c)** Performance is calculated as the average payoff rate per supergame in a round-robin tournament in which each strategy is matched with each other strategy from the corresponding stage 1000 times. The length of each interaction is determining by the continuation probability from the experiment, $\delta = 0.95$, with the same 1000 draws for each pair of strategies. Payoff rate is the average earning per period within a supergame. Clusters are identified by color. Performance in LR1 is opaque. Arrows denote change in performance from LR1 to LR2.

The fourth result is regarding the evolution of strategies between the two locked-response stages:

Result 4 *Strategies are becoming less random over time.*

Using the data presented in Figure 3, we identify three key observation regarding the evolution of strategies. First, subjects were less likely to make changes to pure strategies and more likely to make changes to mixed-strategies. This can be seen in panel (a) as only 35% of subject with average RRM less than 5 in LR1 changed their strategy, while 85% of subjects with average RRM greater than 5 in LR1 changed their strategy (p-value < 0.01, using Fisher's exact test). Second, of the strategies that were changed, more changed to lower average RRM (25, denoted by green arrows) than higher average RRM (9, denoted by red arrows). Finally, the figure shows that pure-strategies perform better. Panel (a) show that 24 out of top 26 best performing strategies had an average RRM of less than 5. Panel (b) shows that all but one of the top 26 strategies were classified in the pure-strategy clusters GRIM, ALLD, and TFT. The only strategy that wasn't classified in these three clusters was pure STFT, which, somewhat unexpectedly, was the best performing strategy in our simulations.[16] These three observations provide strong evidence that more subjects may start playing these well-performing pure strategies if given more time to learn.

4 Discussion

We build on a novel experimental approach to elicit mixed-strategies in the indefinitely repeated prisoner's dilemma with a high continuation probability ($\delta = 0.95$). Our experimental interface allows subjects to construct memory-1 mixed strategies using a set of "if [input]-then [output]" rules, where the output is a probability. We find that although a majority of subjects use mixed strategies, the majority of behavior is best characterized by three pure strategies: TFT, GRIM, and ALLD. In addition, we find a notable fraction of subjects that use mixed strategies that do not closely match the three pure strategies, and which we refer to as the mixed-tit-for-tat class of strategies. When looking at the evolution of strategies, we find that subjects that used pure strategies in the first locked-response stage continued to use them in the second locked-response stage, while subjects that played mixed-strategies tended to make their strategies less random. Next, we discuss these results in connection to the literature and outline several interesting avenues for future research.

The first main takeaway is that three commonly studied pure strategies do a good job in explaining which strategies subjects construct. In particular TFT, GRIM, and ALLD, account for approximately 60 percent of strategies in our experiment, which is in line with findings in the literature. This result is particularly striking in the context of our experiment because these three

[16] In the Appendix, we provide further investigation of strategy performance. Specifically, Figure A-5 presents evidence on the variability of the performance for each strategy. Notably, STFT and ALLD yield large variability in performance and strong negative relationship with the realized supergame length. While TFT and GRIM generate small variability in performance and weak positive relationship with realized supergame length.

strategies emerge even when the strategy space contains 101^5 strategies. In addition, in a round-robin tournament among all elicited strategies, we find that the the same three pure strategies are among the best performing strategies in the experiment. Furthermore subjects tended to keep playing these strategies once adopted, which suggests that more subjects may start playing these pure strategies if given more time to learn.

The second main takeaway is a clear pattern among the mixed strategies that subjects use. In particular, we find that approximately 30 percent of subjects use mixed strategies in the mixed-tit-for-tat class of strategies. These strategies cooperate with high probability after mutual cooperation, defect with high probability after mutual defection, and randomize otherwise. We considered labeling these strategies as SG (proposed in Breitmoser (2015)) as it has the same trends after CC and DD, and also randomizes after CD and DC. However, we decided against it because SG is defined as having an equal probability of cooperation after CD and DC and the trends in our data suggest that there is a consistent difference in these probabilities in favor of more cooperation after DC than after CD.

There are several promising avenues for future research. First, it would be interesting to test whether the large proportion of pure strategies and the identified type of mixed strategies are observed in treatments with different parameters. In particular, parameters that are studied in Breitmoser (2015) may show more evidence of SG because the SG equilibrium described in that paper is fairly close to GRIM for our experimental parameters. Second, it would be interesting to focus on the learning process with mixed-strategies in more detail. In particular, will the mixed-tit-for-tat strategies go away with experience as suggested by the changes between the two locked-response stages from our experiment? Third, it would be interesting to identify the extent to which subjects actually intend to play longer pure strategies (e.g., TF2T or 2TFT) rather then playing mixed strategies, and vice-versa. Specifically, a design allowing for both memory-2+ strategies and mixed strategies may isolate which strategies subjects use. Finally, future research can use our elicitation approach for evaluating theoretical refinements in repeated game strategies and different strategy estimation procedures.

References

Backhaus, T., and Y. Breitmoser (2018): "God Does Not Play Dice, but Do We?," Discussion paper, CRC TRR 190 Rationality and Competition.

Bigoni, M., M. Casari, A. Skrzypacz, and G. Spagnolo (2015): "Time horizon and cooperation in continuous time," *Econometrica*, 83(2), 587–616.

Bloomfield, R. (1994): "Learning a mixed strategy equilibrium in the laboratory," *Journal of Economic Behavior & Organization*, 25(3), 411–436.

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

Camera, G., M. Casari, and M. Bigoni (2012): "Cooperative strategies in anonymous economies: An experiment," *Games and Economic Behavior*, 75(2), 570–586.

Cason, T., and V.-L. Mui (2017): "Individual versus Group Choices of Repeated Game Strategies: A Strategy Method Approach," *Purdue Working Paper*.

Cason, T. N., D. Friedman, and E. Hopkins (2013): "Cycles and instability in a rock–paper–scissors population game: a continuous time experiment," *Review of Economic Studies*, 81(1), 112–136.

Chmura, T., and W. Güth (2011): "The minority of three-game: An experimental and theoretical analysis," *Games*, 2(3), 333–354.

Cockburn, I., and R. Henderson (1994): "Racing to invest? The dynamics of competition in ethical drug discovery," *Journal of Economics & Management Strategy*, 3(3), 481–519.

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.

——— (2017): "Strategy choice in the infinitely repeated prisoners dilemma," *Working Paper*.

——— (2018): "On the determinants of cooperation in infinitely repeated games: A survey," *Journal of Economic Literature*, 56(1), 60–114.

Ely, J. C., J. Hörner, and W. Olszewski (2005): "Belief-free equilibria in repeated games," *Econometrica*, 73(2), 377–415.

Ely, J. C., and J. Välimäki (2002): "A robust folk theorem for the prisoner's dilemma," *Journal of Economic Theory*, 102(1), 84–105.

Embrey, M., F. Mengel, and R. Peeters (2016): "Eliciting Strategies in Indefinitely Repeated Games of Strategic Substitutes and Complements," *Working Paper*.

Frey, B. J., and D. Dueck (2007): "Clustering by Passing Messages Between Data Points," *Science*, 315(5814), 972–976.

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.

Kahn, L. M. (1993): "Unions and cooperative behavior: The effect of discounting," *Journal of Labor Economics*, 11(4), 680–703.

Maggi, G. (1999): "The role of multilateral institutions in international trade cooperation," *American Economic Review*, 89(1), 190–214.

Mailath, G. J., and L. Samuelson (2006): *Repeated games and reputations: long-run relationships.* Oxford university press.

Noussair, C., and M. Willinger (2011): "Mixed strategies in an unprofitable game: an experiment," Discussion paper, Citeseer.

Ochs, J. (1995): "Games with unique, mixed strategy equilibria: An experimental study," *Games and Economic Behavior*, 10(1), 202–217.

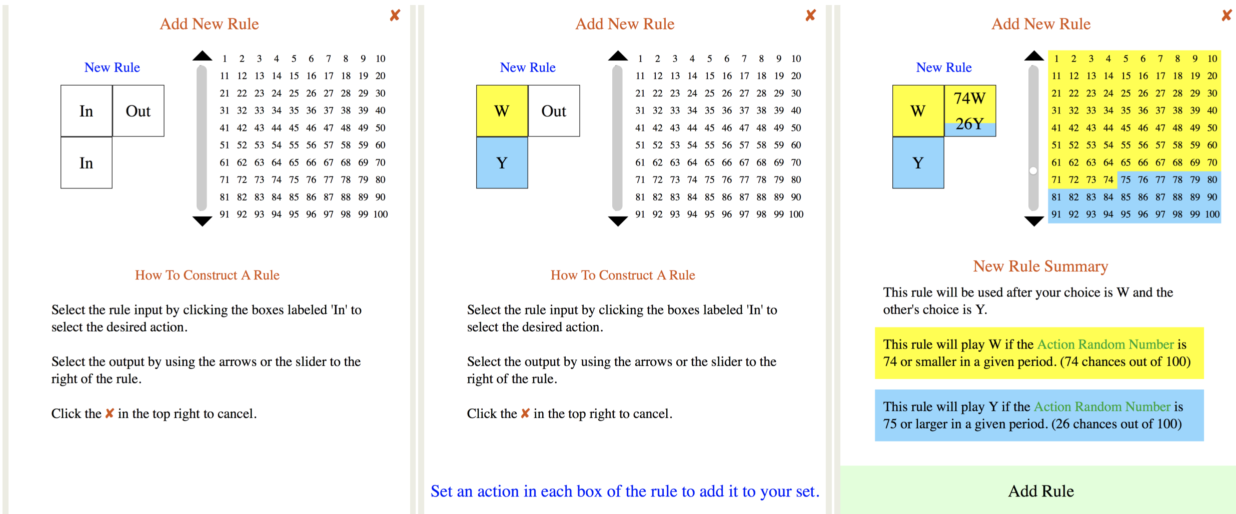
Powell, R. (1993): "Guns, butter, and anarchy," *American Political Science Review*, 87(1), 115–132.

Romero, J., and Y. Rosokha (2018): "Constructing strategies in the indefinitely repeated prisoners dilemma game," *European Economic Review*, 104, 185 – 219.

Shachat, J. M. (2002): "Mixed strategy play and the minimax hypothesis," *Journal of Economic Theory*, 104(1), 189–226.

Appendix A Additional Tables and Figures

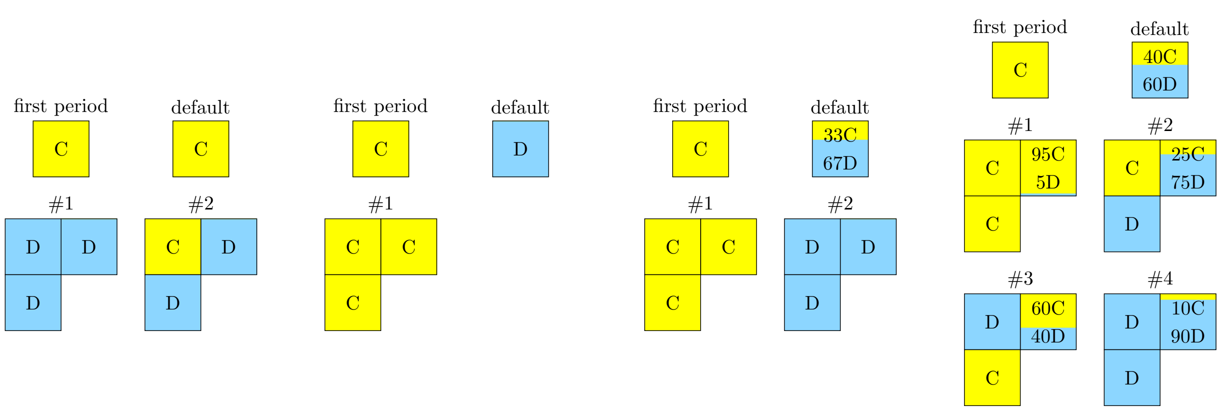
Figure A-1: Rule Constructor Screen-shots.



(a) (b) (c)

Notes: (a) Before any selection has been made; (b) After the input has been set, but before the output has been set; (c) After both input and output has been set. Subjects could make selection regarding inputs and output in any order they choose.

Figure A-2: Examples of Rule Sets.



(a) TFT (b) GT (c) SG (d) mixed-TFT

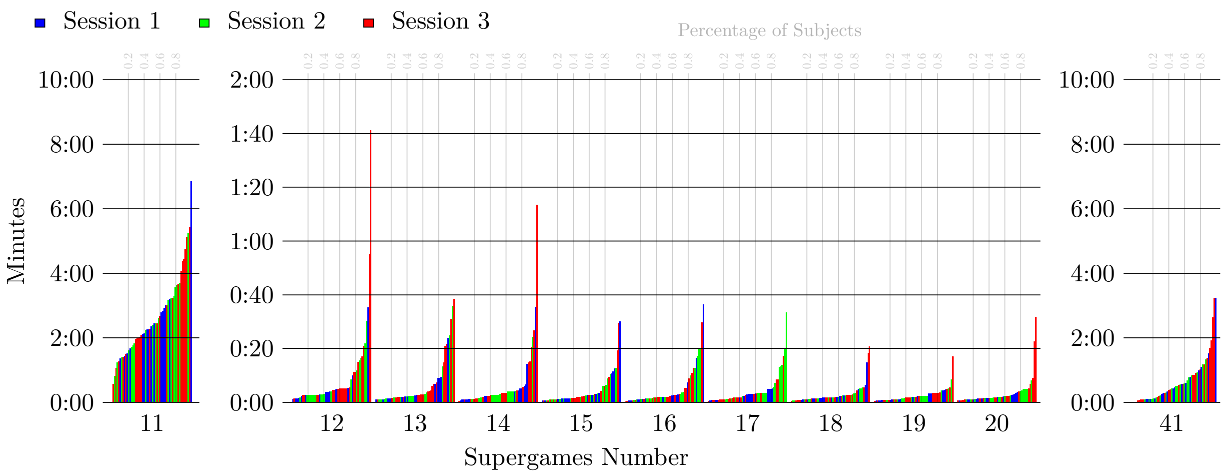
Notes: (a) TFT – Tit-for-Tat; (b) GT – Grim Trigger; (c) SG – Semi-Grim; (d) mixed-TFT – mixed tit-for-tat. Note that there are multiple ways to construct the same strategy.

Table A-1: Supergame Length Realizations.

Supergame Number:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
Realization #1:	5	11	9	2	11	2	12	25	4	57	25	4	20	25	89	3	15	19	42	13	16	2	9	9	21	7	12	8	28	10
Realization #2:	19	25	7	9	2	7	4	54	30	19	24	16	3	15	39	62	14	12	13	9	18	47	11	51	40	4	31	3	4	8

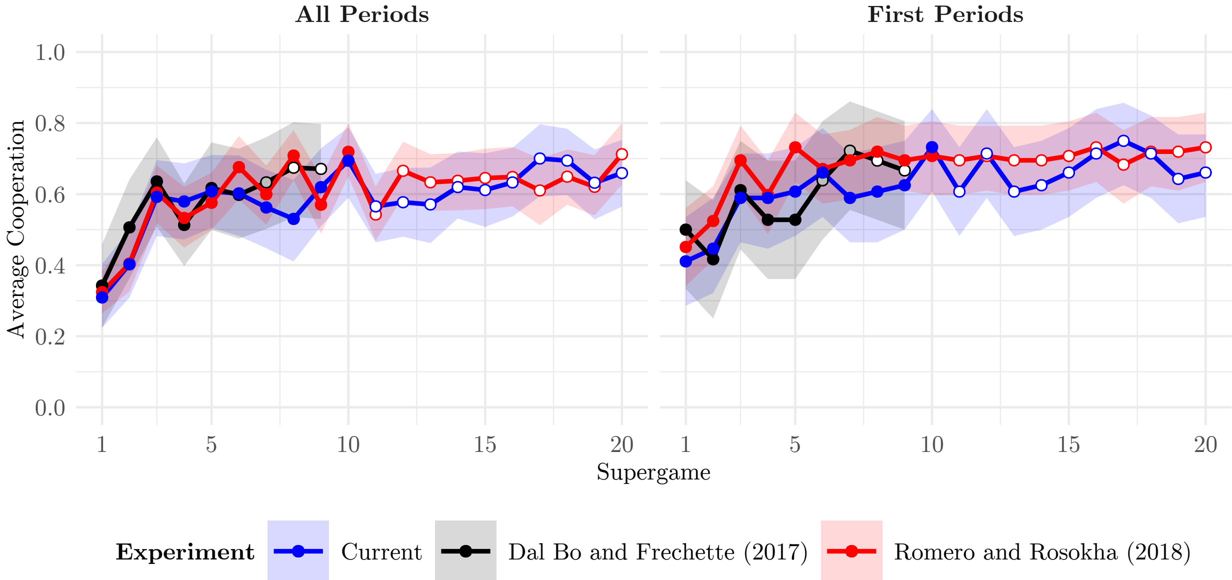
Supergame Number:	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	**60**
Realization #1:	39	5	27	22	23	2	7	2	2	12	71	9	37	45	27	16	1	61	4	12	4	7	69	15	23	14	137	6	11	63
Realization #2:	9	31	22	3	18	34	20	18	43	14	17	2	26	13	23	4	29	7	17	3	4	6	13	28	2	27	9	14	15	79

Figure A-3: Time Before Start Match Button Click.



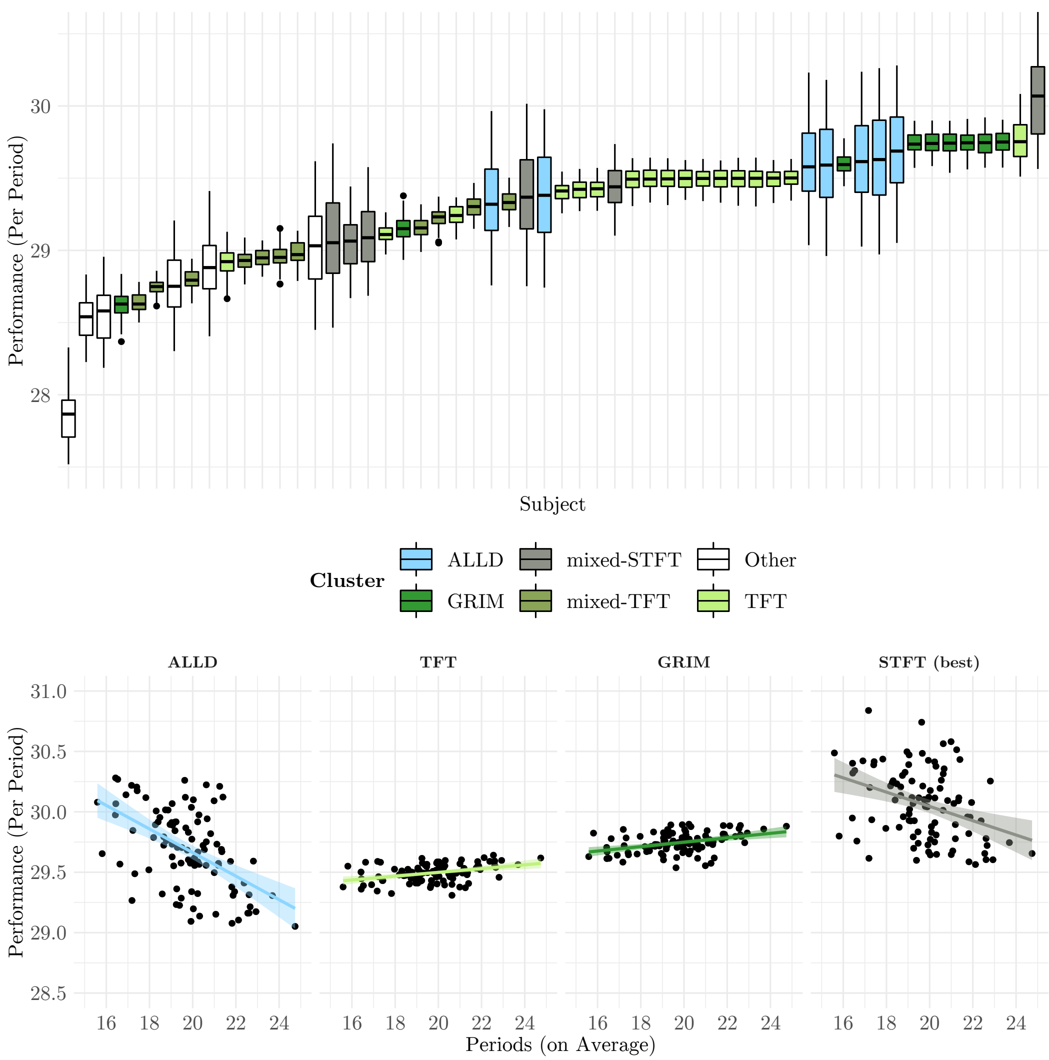
Notes: Cumulative distribution of times at which subjects clicked start match button. They had up to 10 minutes before the first supergame in the non-binding stage (supergame 11) to construct the initial rule set. Once a subject clicked 'start match' button, they were still able to make changes to their rule set until everyone clicked 'start match' button. Subject had up to 2 minutes before each of the other supergames in the non-binding stage. Subjects had up to 10 minutes before the second locked response stage.

Figure A-4: Cooperation Comparison.



Notes: Solid points represent direct-response stage. White points represent non-binding stage. 95% bootstrapped confidence intervals are superimposed. Dal Bo and Frechette (2017) run two sessions with different numbers of supergames in each stage. Gray points for (supergames 6 and 7) represent one session being in the direct-response stage while the other being in the non-binding stage.

Figure A-5: Strategy Performance Variability.



Notes: We ran 100 round-robin tournaments in which each strategy from the second locked-response stage was matched with each other strategy from the population for 100 supergames of random duration (supergame lengths were the same for each pair within a tournament, but different across tournaments). **Top**: Variability in strategy performance across 100 tournaments. **Bottom**: Each point denotes the strategy performance (the average payoff rate per supergame) in one tournament and the realized average supergame length in the same tournament. Payoff rate is calculated as the average earning per period within a supergame. Clusters are identified by color.