Team versus Individual Play in Finitely Repeated Prisoner Dilemma Games*

John H. Kagel
Department of Economics
Ohio State University

Peter McGee
Department of Economics
National University of Singapore
3/10/2014

Abstract

In finitely repeated prisoner dilemma games, two-person teams start out with significantly less cooperation than individuals, consistent with results reported in the social psychology literature. However, this quickly gives way to teams cooperating significantly more than individuals. Team dialogues show significant discrepancies between beliefs and those underlying leading "rational" cooperation models. Cooperation is sustained by its higher payoff in conjunction with limited unraveling due to teams (and by implication individuals) failure to account for others planning to defect earlier based on past experience, much as they do, which is consistent with boundedly rational thinking.

Key words: finitely repeated prisoner dilemma games, team versus individual play, boundedly rational thinking.

JEL classification: D03, C92, C73

*We have benefited from discussions with Guillaume Fréchette, P. J. Healy, Chester Insko, Jim Peck and Tim Wildschut, as well as comments at a brown bag talk at Ohio State University. Valuable research assistance was provided by Xi Qu, Matt Jones, Dimitry Mezhvinsky, and Andrzej Baranski. This research has been partially supported by National Science Foundation grant SES-1226460. Opinions, findings, conclusions or recommendations offered here are those of the authors and do not necessarily reflect the views of the National Science Foundation.

We report the results of an experiment comparing the behavior of two person teams in a finitely repeated prisoner's dilemma game (FRPD) with individuals playing the same game. There are two motivations for the experiment. First, given the high frequency with which economic decisions are made by teams as opposed to individuals, there is interest in what, if any, differences will emerge between the two. Second, and of equal interest, is to use the team dialogues to sort out between different models for, and to better understand, the pattern of play typically reported in these games – increased cooperation developing over time in early stage games followed by defection as the end game draws near and very limited unraveling.

The experiment sits at the intersection of two strands of research – social psychology experiments that compare team versus individual play in FRPD games, and economic experiments designed to investigate models aimed at organizing typical patterns of play in FRPD games. The main finding from the social psychology experiments is that teams are significantly less cooperative than individuals, a result referred to as the "discontinuity effect" (see Wildschut et al., 2003 and Wildschut and Insko, 2007, for surveys of the literature). This result is reported in two, three and four person teams and is accentuated when agents have face-to-face for discussions before deciding what actions to take. Explanations for the discontinuity effect are offered in terms of greater fear, distrust and greed between inter-group than inter-individual relations, or as group discussions facilitating more rational comprehension of the situation, thereby leading to superior backward induction and less cooperative play. We expand on these explanations when reviewing the social psychology literature on the discontinuity effect.

Our experimental results, which do not involve communication with rivals, are consistent with the discontinuity effect in that there is significantly less cooperation for teams in the first super-game played, with the team dialogues showing that this is largely the result of "safety" considerations (namely fear of the other team defecting and getting the "sucker" payoff). However, after this first-super-game, teams cooperate at the same or higher levels than individuals, with significantly higher levels of cooperation in later super-games. There is no corresponding social psychology data that we are aware of against which to compare this last result, as the typical experiment involves a single super-game.[1] Team dialogues indicate that

[1] Based on personal communication with Professors Tim Wildschut and Chester Insko.

early concerns with 'safety" give way to a willingness to take some risks in order to earn the higher profits from cooperation.

Economic experiments have focused on the Kreps et al. (1982) model of behavior in FRPD games.[2] This elegant model shows that if perfectly rational agents believe that there are sufficient numbers of conditionally cooperative types in the population ("crazy" types), it is in their best interest in early stage games to play cooperatively, only to defect as the end game draws near. The percentage of conditionally cooperative types needed to support this model can be surprisingly small; in fact, the model does not even require that there actually be any conditionally cooperative types in the population, as actions are driven by beliefs (Reny, 1992). This argument serves to rationalize, at least qualitatively, the typical pattern of play in FRPD experiments after subjects have gained some experience. There have been a large number of experiments investigating more detailed predictions of this model with mixed results, which are briefly discussed in the next section of the paper.

Our experimental results show a number of similarities in outcomes between teams and individuals; e.g., over seventy percent of the time when defecting earlier than the last time they were up on a cooperative path, both teams and individuals defect one period earlier, and regression results show that the factors impacting round one cooperation rates are quite similar. Yet teams are clearly more "rational" than individuals as they cooperate significantly less often in the last stage game than individuals do. These differences are large enough that teams satisfy the "truth wins" norm (Lorge and Solomon, 1955), namely that teams are as, or more rational, than the most rational individual in the team. While it is clear from surveys of the literature that teams are more rational than individuals in a variety of games of interest to economists (see Charness and Sutter, 2012 and Kugler et al., 2012), satisfying the truth win's norm has typically not been looked at in the economics literature, and it is rarely satisfied in the psychology literature (Davis, 1992).

The team dialogues allow us to directly investigate agents' beliefs, which are critical to economic models designed to explain the characteristic pattern of play. These show that fear of being defected on give way to taking a chance on cooperation based on its higher payoffs. At the same time as sustained cooperation rates are increasing over super-games, there is increasing

[2] This stands in contrast to what happens if all agents are rational own income maximizing types, and believe that all other agents are rational own income maximizing types: then the game should completely unravel with no cooperation between agents. This should hold even under pre-play communication between agents.

awareness that the other team will certainly defect prior to the last stage game. This leads to some unraveling which is quite incomplete as there is very little, if any, consideration of the fact that other agents are learning to defect earlier as well. As such it is clear that "rational" models of cooperation based on "crazy" types is not driving behavior, but rather a number of elements of bounded rationality. We argue that individuals suffer from the same breakdown of common knowledge of rationality, but with even more limited insight as they as they start out defecting later, and unravel more slowly over time than teams do.

The remainder of the paper is organized as follows: Section I reviews prior research on FRPD games that helped to establish the issues that our empirical analysis deals with. Section II outlines our experimental design and procedures. Section III reports the experimental results in relationship to the issues raised in Section I. This has two distinct parts – comparing individuals with team play and using the team dialogues to gain insight into subjects' beliefs driving their behavior. Section IV briefly summarizes our main results and conclusions.

I. Prior Research:

There has been much work done on FRPD games in both the economics and social psychology literature. The goal in this section is not to provide an exhaustive review of the literature, but to summarize results from papers most closely related to the work reported here. Within economics, the major puzzle is why these finitely repeated games do not completely unravel, or at least do so over time with experience. Within the social psychology literature much of the focus has been on the "discontinuity effect", the fact that teams tend to cooperate less than individuals.

In the social psychology literature, the evidence for more limited cooperation is strongest when there is communication between rivals along with inter-group discussions (Wildschut et al., 2003; Wildschut and Insko, 2007).[3] Typical procedures here are to first have within-team discussions, followed by discussions between representatives of each team, followed by teams independently deciding on what to do, with corresponding procedures for individuals.[4] A number of clever experimental designs have been employed to try and tease out the reasons why teams are more competitive. Within that literature, there are two different perspectives: One is that intergroup relations are characterized by greater fear and greed than inter-individual

[3] Note the discontinuity effect is by no means limited to experimental designs involving discussions between players.

[4] Much of this research has involved financially incentivized agents.

relations leading to less cooperative play. The second perspective is that group discussion facilitates rational comprehension of the forces at work in mixed motive situations like FRPD games, with the greater rational comprehension favoring greater backward induction (hence less cooperation) on the part of teams.

One of the shortcomings of this literature, from an economist's perspective, is that these experiments have typically involved a single super-game between a pair of agents, as opposed to the typical economic experiment where agents engage in a number of super-games, being rematched following each super-game.[5] Among other things this means that there has been no investigation of whether the greater fear and greed on the part of teams will persist over time, as opposed to team discussions recognizing the benefits of mutual cooperation in early round play. By randomly rematching agents between super-games within a given experimental session, we are able to investigate this issue, as well as whether or not cooperation unravels faster over time for teams as opposed to individuals.

Much of the economics literature has focused on investigating different formal models rationalizing the typical pattern of play reported in FRPD games: an initial period of cooperation followed by cooperation breaking down near the end of each super-game (Selten and Stoecker, 1986, Andreoni and Miller, 1993). With fully rational agents, the standard backward induction argument for finitely repeated games results in defection in every game as the unique dominant-strategy equilibrium. The favorite alternative to this model is the Kreps et al (1982) reputation model. In this model, if there is incomplete information about the types of players one is likely to face, with a high enough probability that some of these agents will be committed tit-for-tat (TFT) players, then cooperation in early plays of the game is consistent with fully rational behavior, along with defection as the end game draws near. The model is static, failing to account for the typical pattern of greater early round cooperation over repeated super-games, followed by some unraveling (earlier defection) in later super-games.

Early experiments provided qualitative support for the model: Andreoni and Miller (1993) compared FRPD games with one-shot PD games, reporting substantially more cooperation in early stage play in the FRPD games, and close to the same level of end game cooperation as in the one-shot games, consistent with agents believing there are "altruistic-types"

[5] The one-shot super-game nature of most of this research is not obvious from reading the surveys on the discontinuity effect.

in the population. In addition, FRPD games in which there was a 50% chance of playing against a computer playing TFT (where the TFT strategy was announced and explained to subjects) resulted in cooperation being sustained at very high levels for substantially more early stage games than with all human competitors.[6]

Subsequent experiments have not provided such strong support for the Kreps et al. model. Cooper et al. (1996) report results for a 10-period FRPD game which exhibits the typical pattern of early cooperation followed by defection, along with higher cooperation rates in early plays of the game than in a one-shot game. While this aggregate pattern of play is qualitatively consistent with the Kreps et al. model, using more detailed data they report contrary evidence as (i) at times cooperative play follows non-cooperative play by the same player or his opponent, which should not occur, and (ii) the high levels of early cooperation observed require substantially higher levels of conditionally cooperative types than found in end period play.[7] Cooper at al. note that the inability of their data to fit the Kreps et al. model better could reflect the limited type of "irrationality" underlying the original model, as a substantially wider variety of equilibria can result from alternative "irrational" types (as in the infinitely repeated game literature; Fudenberg and Maskin, 1986) and /or behavior may be driven in part by mistakes on the part of players.

Cox et al. (2012) investigate the Kreps et al. model within the context of a finitely repeated *sequential* PD game, where second movers decide what to do after seeing what first movers did.[8] Their experiment involved two blocks of 5 super-games of 10 rounds each. Unbeknown to subjects, their choices in Block 1 were revealed to players in Block 2 under one of two information conditions: (1) One-sided information where only the first mover's history is revealed and (2) Two-sided information where both players history is revealed. If the Kreps et al. reputation building model holds, then player 1's matched with a player 2 who defects in later rounds of block 1, reveals that their opponent is not a committed TFT player, which should destroy early round cooperation. The aggregate data is clearly inconsistent with this prediction as the frequency of first mover cooperation is significantly greater with two-sided information

[6] Also see Camerer and Weigelt (1988) who report strong support for the Kreps et al. reputation building model in the context of a borrower-lender game.
[7] Similarly, Jung et al. (1994) look at the Kreps et al. reputation building model for limit-entry pricing in the context of the chain-store paradox, finding substantially weaker support for the model than Camerer and Weigelt report.
[8] In a related experiment, Reuben and Suetens (2012) look at a finitely repeated sequential PD game with an uncertain end point. Using the strategy method, first movers can condition their actions on whether or not the super-game will end, with second movers able to condition their actions on this and the first mover's actions.

compared to one-sided information, and is greater than first movers' cooperation in Block 1.[9] Cox et al. rationalize these findings by relaxing the requirement that a player's prior beliefs be consistent with their opponent's best response.

Selten and Stoecker (1986) focus on learning in a simultaneous move FRPD game. Their interest in learning is motivated by the fact that levels of cooperation and defection change in consistent ways over time. Their paper focuses on end game behavior, employing a Markov learning model where subjects change their intention to deviate from cooperation depending on their experience in the previous super-game. Subjects were asked to write down reasons for each period's decision, with these descriptions, in conjunction with observed patterns of play, used to determine the period in which subjects intended to defect. In their model, defections between super-games does not occur, or shifts one period earlier or later than in the previous super-game. Defection occurs earlier if their opponent deviated earlier than they intended to, or deviated in the same period as they intended to, with increased likelihood of defecting earlier if their opponent deviated before they did. Defection is likely to occur one period later if a player defected before their opponent did in the previous super-game. They find strong support at the individual subject level for their learning model in later super-games.

The Selten and Stoecker model is essentially one of bounded rationality. There are other bounded rationality models applied to FRPD games. Bereby-Meyer and Roth (2006) report an experiment in which the speed of adjustment to the mature pattern of play reported in noisy FRPD games is captured by reinforcement learning models from psychology. Jeheil (2005) develops a boundedly rational model which essentially has agents establishing equivalence classes across rounds for when their opponent is likely to defect and best responding to this. In this model, agents always defect in the last stage game and, like the Selten and Stoecker model there is no characterization of the underlying behavioral forces that drive the typical learning process. Neyman (1985) develops a model of cooperation in FRPD games under the assumption that there are bounds to the complexity of the strategies that players can use.

Our experiment takes place at the intersection of the social psychology and economics literatures. To our knowledge it is the first study of team play in finitely repeated PD games with a number of rematches following each super-game, which can help to determine if the

[9] There are also significantly higher levels of conditional cooperation on the part of second movers with two sided information.

"discontinuity effect" persists with experience. Are the beliefs underlying team play, revealed through team dialogues, at all consistent with the Kreps et al. model or better approximated by a model of bounded rationality? In addition, we can use the team dialogues to better understand how cooperation tends to grow over time and, while unraveling somewhat with experience, is such a slow and limited process. Further, from the point of view of whether teams are more "rational" than individuals, we compare end period levels of cooperation between the two, and whether what unraveling does occur is faster for teams.

II. Experimental Design and Procedures:

Subjects played a 10 stage, simultaneous move FRPD with stage game payoffs reported in Figure 1. Payoffs were denominated in experimental currency units (ECUs) which were converted into dollars at the rate of $1 = 250 ECUs. Payoffs were computed over all plays of all the super-games and paid in cash at the end of an experimental session along with a $6.00 participation fee. Each member of a team received his team's payoff.

[Insert Figure 1 here]

In the teams treatment subjects were randomly matched with a partner at the beginning of a session, with partners remaining the same throughout the session. Teams played against teams, and individuals played against individuals. In what follows we will refer to agents playing a game, with agents either two person teams or an individual. Following each FRPD game, agents were randomly re-matched under the restriction that no two agents would be re-matched in consecutive super-games. Teams in sessions1-3 each played seven FRPD super-games. This was increased to nine and ten FRPD super-games in sessions 4 and 5, as it was clear there was sufficient time to play the extra games. All five individual subject sessions played ten FRPD games. Each session had between 8 and 12 individual subjects/teams for a total of 52 individual subjects and 51 teams. The discrepancy in the number of agents across treatments is due to the use of a student assistant to ensure an even number of teams.[10] In both treatments agents were given a range of possible super-games they would play, inclusive of the number actually played.

Teams had 3 minutes to discuss and make their choices in the first two rounds of each super-game. This was reduced to 1.5 minutes after that. Default options if time ran out without

[10] The assistant informed his teammate that he was one of the experimenters and would agree to whatever his partner did. He also asked his partner to write out any thoughts he/she had about the game in the chat box as they played. Data for this team is dropped except as needed to complete play when paired with another team.

a coordinated choice are enumerated in the instructions, which can be found in the online appendix.[11] Similar time limits were imposed for individual play, but these were never binding.

Following the end of each stage game agents had up to 30 seconds to view the results before moving on to the next stage game. Following the last stage game, agents were notified that their match had ended and that they would start another match with another (randomly chosen) agent. Neutral language was used throughout; e.g. agents chose between option A or option B in each stage game, and were told they would be "paired with the **same** other team for a set of 10 repeated choices."

III. Experimental Results:

We report the results in two parts, first comparing patterns of play between individuals and teams, making use of the team chats as needed. We then analyze the team chats in order to better understand the beliefs underlying the behavior reported and to sort out between explanations for the behavior. The analysis is limited to the seven super-games common to all sessions.

III .1 Comparing patterns of play between individuals and teams

Figure 2 reports average levels of cooperation for teams and individuals over the seven super-games. The data exhibits the usual pattern in both cases with cooperation rates at their peak in the early stage games followed by a rather precipitous drop as the end stage draws near.

Table 1 reports average stage one cooperation rates for each of the super-games along with z-statistics for differences between the two cases. The focus is on the stage one cooperation rates as cooperation in later rounds is very much dependent on what happens in the first stage game which creates complicated interdependencies that are difficult to account for. Further, once two or more rounds have passed in which one agent has defected, in the overwhelming number of cases both agents defect for the remainder of the game. Average stage one cooperation rates are significantly higher for individuals compared to teams for the first super-game. However, by the second super-game the rates are essentially the same, with teams having higher average cooperation rates in the remaining super-games, with these differences statistically significant in super-games 5 and 6.

[Insert Figure 2 and Table 1 here]

[11] Overall, 97.5% of all team choices involved active coordination between teammates on choices made. The appendix can be found at http://sites.google.com/site/econpjmcgee/AppendixKM.pdf.

Appendix A1 reports the results of a probit investigating the factors behind cooperation in the first round of play across super-games, employing variables shown to impact first round play in infinitely repeated super-games (Dál Bo and Fréchette, 2012). Key results are that (i) cooperation in round 1 of the first super-game measures an inherent tendency to cooperate that carries over to later games, (ii) agents are more likely to cooperate if the person they were paired with in the previous super-game cooperated, and (iii) the main effect for a team dummy shows higher overall rates of cooperation ($p < 0.10$), other things equal. The first two characteristics hold equally for teams and individuals; i.e., interaction effects between these two variables and teams are not statistically significant at conventional levels.

Conclusion 1: Consistent with the discontinuity effect reported in the social psychology literature teams are less cooperative than individuals in the first super-game. However, they are as, or more, cooperative than individuals in later super-games, so that overall teams are more cooperative than individuals.

Fully rational, own income maximizing agents should never cooperate in the last period of a super-game. In contrast to this, there is some cooperation in end round play for both teams (9.8%) and individuals (26.9%), with this difference significant at the 5% level. [12] Cooperation in the last round of play is sometimes treated as evidence for altruism in the population, typically reciprocal altruists committed to TFT play (Andreoni and Miller, 1993; Cox et al., 2012). However, the team chats suggest that mistakes, confusion or naiveté account for all of the cooperation reported. For example, one team in the next to last stage game defected on the grounds that they would earn the higher payoff, noting that if they did so "…we get 175? we won't ever play them again." But in round 10, one member of the team mistakenly cooperated, with no time to correct the mistake, and the computer selected that player's choice. Or to take another case: After choosing to defect in 8 out of 9 rounds, one team chose to cooperate in the end game "just for the hell of it". That mistakes, confusion or naiveté account for most of the end round cooperation is also supported by the fact that over half of these outcomes occur by super-game two for both teams and individuals.[13]

[12] These percentages are based on cooperating one or more times in the last period of play, with agents cooperating more than one time counted once in the data. No team cooperated more than once, with two individuals cooperating 3 and 2 times respectively.

[13] Also note that for all the agents who cooperated in round 10, there is one other super-game in which that agent fails to act as a committed TFT player, either because they defected first, or defected simultaneously, following a continuous string of cooperative play, or did so prior to a string of cooperative choices.

Given that this end period cooperation is driven by mistakes, confusion or naiveté, teams should have fewer mistakes, which indeed they do. Further, one would expect that teams would meet or beat the truth wins (TW) norm (Lorge and Salomon, 1955) in this case. The TW norm holds that for problems have a clear, correct answer which can be easily explained to one's partner, a team should do as well or better than the best member of that team, as she should be able to explain the solution to her teammate.[14] This can be investigated with a simulation in which individuals cooperate in the end game at the level reported in the population, and determining the frequency with which two randomly selected individuals would cooperate when one or both of them, "solved" the problem as individuals. This occurs on average for between 5.9% and 7.8% for teams, with the observed rate (9.8%) well within the 90% confidence interval for the TW norm.[15]

Conclusion 2: There is significantly less end game cooperative play for teams versus individuals. This is indicative of greater rationality and/or clarity of thought on the part of teams. Further, team chats indicate that this cooperation was a result of mistakes, confusion, or naiveté, which no doubt holds for a number of individuals as well. Simulations show that teams' fall well within the 90% confidence interval of the truth win's norm, which should apply to this situation.

Teams unravel a bit more and faster than individuals. To measure this, we determined the round in which an agent first defects, conditional on being up on a cooperative path at the start of a super-game. The latter is defined as sustained cooperation over rounds 1-4, typically involving both agents cooperating in each of rounds 1-4.[16] Table 2 reports the average number of defections in each super-game along with the round in which the defection occurred. In the first super-game, the median round in which defections occurred was 10 for individuals and 9 for teams ($p > 0.10$, Mann-Whitney test). For both teams and individuals, there is slow, and far from complete, unraveling across super-games, with the median for teams always one step ahead of individuals, until the last super-game where it is two steps ahead (round 7 versus round 9; $p < 0.01$, Mann-Whitney test). This greater unraveling for teams can be attributed to them having

[14] Note that the psychology research on this issue shows that teams rarely meet, no less beat, the TW norm (Davis, 1992). Investigations of team versus individual behavior in economic contexts rarely even address this question, in part because in many cases the insight needed to solve the problem is sufficiently complicated that it would be quite difficult to explain the solution to one's partner(s).

[15] The simulation consisted of samples of 51 teams, drawn from 52 individuals (14 cooperators; the rest defectors) with replacement and repeating the simulation 250,000 times. Simulated teams were counted as cooperating when the two individuals drawn both cooperated.

[16] The exceptions to this criterion are discussed below.

more experience with defection as they get up on a cooperative path significantly more often than individuals, and/or that they are better able to anticipate that others are learning to defect earlier in much the same way they do. However, as the team chats show, there is very little accounting for other teams learning from the similar experience and when they do, following through on the logical implications is quite limited.

[Insert Table 2 here]

We also looked at the extent to which agents responded to defection *between* super-games in which they were up on a cooperative path in rounds 1-4. Table 3 reports these results. There are three categories for what agents were doing when the defection occurred (at the top of the table): (1) an agent was cooperating when he was defected on (Were Cooperating), (2) both agents chose to defect in the same round (Both Defected), or (3) an agent defected on his own while the other agent was cooperating (Unilateral Defection). The percentages show how agents responded the next time they were up on a cooperative path – dropping at an earlier round (Earlier), dropping in the same round (Same), dropping in a later round (Later), or dropping in the same or later rounds (Same or Later). The latter reflects that observations are censored when an agent was cooperating, but was defected on one round earlier than the last time he was up on a cooperative path.[17]

[Insert Table 3 here]

Agents were most likely to defect earlier than the last time they were up on a cooperative path when they defected simultaneously with the agent they had been paired with, with teams holding a slight edge in this regard. Close to three quarters of all these defections occurred one round earlier than the last time they were up on a cooperative path: 72.2% and 73.3% for teams and individuals, respectively.[18]

Agents were least likely to defect earlier when they got the sucker playoff the last time they were up on a cooperative path. One might be concerned that this response is the result of what happened in the intervening super-games in which agents failed to get up on a cooperative path. However, these super-games would have, by definition, consisted primarily of non-

[17] There are two cases to consider: If an agent was defected on 2 or more rounds earlier, we don't know what they would have done. These (few) observations are excluded from the calculations. However, if an agent was defected on 1 round earlier than the last time they were up on a cooperative path, it must be the case that they were planning to defect in the same or a later round.

[18] There was one instance of dropping 3 rounds earlier for one team.

cooperative play. And it is hard to see how this experience would promote defecting later than before. As such we suspect that this strange behavior smacks of agent-specific slow learning.

Agents were most likely to drop in the same round or later when they had unilaterally defected the last time they were up on the cooperative path. Looking at the team chats suggests this was primarily motivated by the fact that once they had defected they were locked into mutual defection, with its lower payoffs, for the remainder of the super-game.[19]

Conclusion 3: Conditional on being up on a cooperative path, there were minimal differences between teams and individuals in defection patterns based on past experience,. The fact that teams unraveled more than individuals by the last super-game is largely attributable to teams defecting earlier in the first super-game, as well as teams having more experience being up on a cooperative path with its opportunities for defection. The fact that when defecting earlier agents typically defected one period earlier suggests limited accounting for other agents learning in much the same way they were, a conclusion that is supported by the team dialogues.

The defection data reported in Table 2 is consistent with Selten and Stoecker's (1986) learning model, as the preponderance of defections occurred one period earlier, with defection occurring in earlier rounds with experience. However, Table 3 shows substantial deviations from the predicted pattern of defection. Selten and Stoecker predict that when an agent is defected on while cooperating, or both agents defect at the same time, that agent is more likely to defect earlier the next opportunity they get, with this likelihood the same or higher when they get the "sucker" payoff ("Were Cooperating" in terms of Table 3). But the data for both teams and individuals shows a different pattern, with a substantially higher frequency of defecting earlier when "Both Defected".[20] What is consistent with the Selten and Stoecker model is that when an agent defects earlier than his rival in the previous super-game ("Unilateral Defection" in Table 3), he is likely to defect in the same period or later at their next opportunity, as the sum of these categories (Same, Later, and Same or later), accounts for close to 80% or more of these responses.

This analysis of defection is conditional on agents being up on a cooperative path. For these purposes we initially defined "up on a cooperative path" as four or more rounds, beginning

[19] From one representative chat: "my only thing with moving it up from round 9 is that, like i said, once you change your selection to B, you're basically stuck on B. there isn't a good different choice."

[20] Using Fisher's exact (two-tailed) test these differences are significant at the 5% level for both teams and individuals. However, these results are only suggestive since our data (i) has repeated measures for the same agent and (ii) occur much earlier in the learning process than those considered by Selten and Stoecker.

with round 1, in which both players cooperated.[21] This is by far the most consistent pattern observed and will be referred to as pattern P_{CC} (see Table 4). However, we identified a number of dialogues in which teams proposed to defect in round 1 followed by cooperation in round 2 if the team they were paired with cooperated in round 1. For example, here is a team discussing what they planned to do in the next super-game:

> 14: you want to do B (defect) again ?
> 9: it's a new team
> 9: i dont know
> 9: but to be safe
> 9: better go with b right?
> 14: i think so, yes.
> 9: go with b first and see what the other team pick for the first round
> 14: if they choose A (cooperate)... that means they want to be nice... so round 2 we'll choose A to apologize

These teams were fully aware that they would more than likely face punishment in round 2, but were planning to cooperate in the hope that their reverting to cooperation would signal to their rival that they wanted to cooperate.[22] Early on this strategy was indeed effective in generating mutual cooperation in rounds 3 and 4 (and beyond), so these cases were also classified as being up on a cooperative path (pattern P_{DC} in Table 4—defection followed by cooperation). In a handful of these potential P_{DC} cases, the team that cooperated in round 1 did not punish in round 2, so that there was mutual cooperation in rounds 2-4.[23] This pattern is also classified as being up on a cooperative path (pattern P_{DC*} in Table 4).

Table 4 reports which of these three patterns was observed the *first* time an agent got up on a cooperative path, as well as the frequency with which these patterns held after that.[24] The patterns P_{DC} and P_{DC*} were relatively common for teams the first time they were up on a cooperative path. However, both patterns largely vanish after that. The data in Table 4 also reflect the higher frequency of cooperation for teams after the first super-game.

[Insert Table 4 here]

[21] The choice of four rounds here is, admittedly, arbitrary but seems natural under the circumstances and corresponds to the number of rounds employed in Selten and Stoecker (1986).
[22] These teams were not naïve as they were prepared not to cooperate in round 4 if their rival failed to get the signal.
[23] The team initially cooperating in this case had a conscious strategy of giving their rival two shots at cooperating before defecting.
[24] Note there is some overlap in the initial frequency with which teams are counted in Table 3 because an agent who first cooperated in super-game *t* might be paired with an agent who first cooperated in a later period. A total of 10 teams and 13 individuals never got up any of these three cooperative paths.

III.2 Team Dialogues in Relationship to Behavior

Table 5 reports the coding categories for team dialogues. There are three broad categories with a number of sub-categories. The broad categories were coded conditional on whether a team was cooperating or not cooperating, along with a general category for coding regardless of whether the team was cooperating or not. Coders were told they could assign multiple codes to the same round; e.g, code C1 and C2 in the same round. Two economics graduate students coded the dialogues. Categories were initially established by the authors after reading a sample of the dialogues. The coders then independently coded a single (common) session, after which they met with one of the authors in order to refine their common understanding of the categories. They then independently coded the rest of the sessions, after which there was a meeting to reconcile obvious discrepancies between the two coders. Across all sessions, the coders were in agreement 76% of the time.[25] In the analysis that follows, when conditioning on a coding category, unless otherwise stated, it is counted if either of the two coders assigned the code in question.

The goal behind the coding is to better understand the beliefs and strategies underlying teams' actions. Our assumption is that these beliefs and ideas are reflective of what is going on with individuals as well. This assumption is based on teams and individuals having very similar patterns of behavior, as have been reported on above. What differs between the two has to do with teams having greater clarity, and earlier insights, into "rational" behavior as might be expected from having two agents deciding cooperatively what actions to take. There is also no doubt that the frequency with which certain beliefs are coded contain some errors, partly reflective of the fact that the between coder agreement rate is not 100%. But to the extent that agents' beliefs are central to understanding their behavior, the team chats provide a natural way of tapping into these beliefs.

[Insert Table 5 here]

One key factor we wanted to identify was the basis for teams' decisions to cooperate or defect in the first super-game. One branch of the social psychology literature on the discontinuity effect attributes the high defection rate to teams opting for the "safest choice" (code D1 in Table 5). That is to guarantee the payoff of 75 as opposed to the possibility of cooperating and getting the sucker payoff of 5. For round 1 of the first super-game, 91.7% (22/24) of the

[25] The same code assigned to a different round of the same super-game was counted as a disagreement.

defecting teams were assigned code D1.[26] The following provides an example of one of these dialogues:[27]

> 16: Pick B (defect) every time, yes?
> 1: what do you want to go with?
> 16: If we choose A (cooperate) we get 105 or 5
> 16: if we pick B we get 175 or 75
> 16: seems to me B is the choice in every situation
> …..
> 16: if we pick B every time our minimum amount of money is 21 dollars
> 16: I don't want to jepordize that minimum with some 5 point takes

Of the teams coded as cooperating in round one of super-game one, 70.6% (12/17) were coded as either C1 or C3 – cooperating in order to elicit cooperation with its increased earnings.[28] An example of this is:

> 2: what do you think we should do
> 17: so i say pick a (cooperate)
> 2: ok thats fine. i hope the others arent greedy
> 17: bc that would give us higher payoff average
> 17: if we fall in the A-A zone
> 2: alright im game, lets do it

The increase in team cooperation rates beginning in super-game two is associated with teams noting the advantages of early cooperation in later periods of super-game one or at the beginning of super-game two. An example of this is for a team stuck in mutual defection in super-game one (coded as D5):

> 18: B (defect) agian?
> 18: or do you want to lose money to get them to mutually choose a (cooperate)?
> 6: we'll ride b the rest of the way out this block but i think the best option is to go A the first 2 blocks, see if the other team catches on and if so choose A mutually for the remainder of that block

One unexpected bonus of the coding was to identify the cooperative strategy P_{DC} noted in Table 4. We doubt if we would have ever recognized this strategy without the chats. This is also one reason why we prefer unstructured team talk: it's harder to code compared to structured

[26] Six teams defected with no code recorded for round one.
[27] Note that the dialogues are not edited to correct for grammatical or spelling errors.
[28] The remaining teams who cooperated in round 1 were coded as either discussing strategies in response to a breakdown in cooperation (2 teams, Code C2 in Table 5), with four teams who cooperated assigned no code for round one of super-game one.

communication, but has the potential to identify strategies that would not have been considered *a priori*.[29]

A second objective of the coding was to better understand the factors underlying teams' decisions when to defect, conditional on being up on a cooperative path. Although we had a code X4 for complete unraveling, it was never assigned, as there was never any discussion of anything approaching the full backward induction argument. Code X3 was designed to capture partial unraveling, discussions of whether and when to defect earlier than in previous super-games. A significant insight from these X3 dialogues is that teams typically were myopically best responding to what had happened to them the last time they were up on a cooperative path, with little if any consideration of their rivals thinking along the same lines based on similar experience. For example:

5: ok so next time i think we should try B on turn 9 in the same situation
7: yeah I was thinking about that
5: since the previous two times they had B for the last one anyway
7: right
7: so we'd gain 70 on turn 9

Note the absence of any consideration of the fact that their opponent in the next round might be thinking along the same lines. This failure to think from their opponent's point of view is entirely consistent with the fact that teams defected one round earlier 72.2% of the time.

Teams that defected more than one round earlier than past defections give some limited consideration to opposing teams learning from similar past experience as well. The following example is for a team that in super-game 3 moved to defecting two rounds earlier than in super-game 2.[30] By way of background this team got up on a cooperative path in all seven super-games defecting in round 10 in super-game 1, on the grounds that

14: the last round means that we'll not cooperate any more
9: you think they may be choose B for last round?
14: so B will definetely be a better choice.....
9: OK, good thought

They defect in round 9 in super-game 2 on the grounds that

9: you mean like we start choosing b at round 8?
9: no i mean 9..

[29] In addition, unstructured dialogues are less leading, thereby reducing the potential for demand induced effects.
[30] This particular team got up on a cooperative path in all seven super-games.

9: so we get a 175 and a 75?
14: yes
14: that's better than 105 and 75
14: or (defect) at round 8[31]

Finally in super-game 3

> 14: i think we can start choosing B (defection) at round 7
> ……
> 14: anyway start at round 7 is safe

This involves implicitly acknowledging that other teams were defecting earlier as well, and best responding to these beliefs, but again not fully accounting for other teams learning the same lessons they had. This is not a bad strategy as they defect earlier than their opponents up to super-game 6, at which point another team beats them to the punch.

In addition, there are occasions where teams are explicitly accounting for other teams thinking about defecting early, just like they are. But the iteration goes for just two steps:

> 10: chances are, they either got screwed over or screwed someone over on round 10 of last block which means they'll be thinking they should screw us over in the 9th which is why we whould go with B in the 8th.

Finally, there are occasions where teams identify the round in which they are likely to be defected on but fail to best respond:

> 2: round 8 is usually where we get screwed
> 17: TRUE
> 2: i kinda wanna go B. that way if we do then we still get 75

Which they do in Round 8, failing to best respond by defecting in Round 7.

Conclusion 4: When determining what round to defect in, teams engage in very limited discussions regarding the possibility that other teams were thinking along the same lines they were. This is responsible for the slow and limited unraveling (typically defecting one period ahead) observed in the data. This limited accounting for others behavior is inconsistent with common knowledge of "rationality" or best responding to "crazy" types, but is consistent with boundedly rational thinking.

[31] In this discussion there is a single throw-away line directed at their rival teams strategy: "how early do you think the other team would start to choose B (defection)", with no response to this inquiry.

We coded for dialogues indicating that one or both members of the team had some prior experience with PD games of one sort or another (code X1). This was done regardless of whether the prior experience was from a PD experiment or from classroom instruction. Eleven teams were coded as X1. These teams did not universally cooperate to begin with, or defect particularly early to begin with, when up on a cooperative path. Their past experience was *not* always particularly helpful either, as the following case illustrates:

1: did you hear about gaming theory
16: no, what's that?
1: i guess that is about the same scenario
1: of this experiment
16: oh, ok. i've done this experiment before in sociology with the prisoner's dilemma and its the exact same thing. and youre always supposed to pick B (defect)

This team chose to not cooperate throughout super-games 1-3, turning to cooperate, successfully in super-game 4, and continuing to do so through super-game 7.

More generally, we pose two questions for teams with prior experience: Are these teams more or less cooperative to begin with and is their unraveling process, once up on a cooperative path, materially different from those without any experience? Regarding differences in cooperation rates, teams with prior experience are more cooperative in the first play of the first super-game, 54.5% versus 37.5%, but not significantly so ($p = 0.30$, two-tailed Z-test). With respect to X3 codes, 54.5% of teams with prior experience were coded X3 (discussing defecting earlier than in a previous super-game) versus 50.0% of those with no prior experience. Further, the extent to which those with prior experience unravel is not materially different: The median for the *earliest* round in which a team defected, conditional on being up on a cooperative path, was the same for teams with and without prior experience with PD games (round 7).[32]

We also used the dialogues to determine if teams had a clearly articulated strategy for dealing with their rival's defection when cooperating in *early* plays of a super-game - namely proposing to defect themselves in the next round or two (codes C2 and D5). The following is representative of a C2 dialogue articulated in round 1 of super-game 1:

2: what do you think we should do
17: so i say pick a (cooperate)

[32] This is restricted to cases where a team unilaterally defected or defected simultaneously with the team they were paired with, conditional on being up on the cooperative path. Conditional on being up on a cooperative path one time or more, 3 teams with prior experience (33.3%) were always defected on while cooperating versus 6 teams (18.8%) with no prior experience.

2: ok thats fine. i hope the others arent greedy
17: bc that would give us higher payoff average
17: if we fall in the A-A zone
2: alright im game, lets do it
2: … if they choose b (defect) and we only get five what should we do next time
17: pick B for the next round and see what they choose this time around?
17: and if they pick B again, we stick with B i guess

And a representative D5 dialogue in round 4 of super-game 1:

23: B (defect) again?
7: yeah
23: lame
23: we could be making so much more if we all chose A (cooperate)
23: I say the next block we do 2 rounds of A
23: to see if the next group knows
23: and if not we'll just go back to B?

A large majority of teams (66.7%; 34) were coded as C2 or D5 by the end of super-game 2. This includes four teams coded for C2 or D5 that never managed to get up on a cooperative path.[33] The number of teams coded as C2 or D5 reached 82.4% by the end of super-game 4. Seven teams were never coded as C2 or D5. Of these, two consistently followed a C2 type strategy, with the remaining five never cooperating, along with receiving a number of D1 codes--defection as the safest strategy.[34] Thus, in total, 86.2% of teams were playing, or planning to play, along the lines articulated in C2 and D5, of which all but two had clearly articulated strategies for this. The remaining 7 teams consistently failed to cooperate through super-game seven, always not cooperating. No team was recorded as discussing cooperation regardless of what the other team did, with minimal (though occasional) regrets expressed for defecting with a cooperating partner.

Conclusion 6: Most teams had clearly articulated strategies for defecting in case their opponents failed to reciprocate cooperation by super-game 2. By the end of super-game 4, 86.2% of the teams were following this strategy, with the remaining teams employing an always defect strategy.

Category X5 was designed to speak directly to the Kreps et al. reputation model, coding team's beliefs regarding what their competitors would do in later rounds. In particular, in Kreps

[33] Two of these 4 were strictly coded as D5 and never cooperated.
[34] The remaining two teams never cooperated, but both were coded as D5 in the last super-game.

et al., "rational" cooperators entertain beliefs that a sufficiently large percentage of their rivals are either committed altruists or committed to playing TFT throughout the super-game. This is used to justify building a reputation for cooperation in early plays of the game. We coded four sub-categories under X5: (i) Other team is clearly going to defect, or likely to defect, at some point in later rounds of play, (ii) Other team will cooperate, or is likely to cooperate, as long as we cooperate for *all* plays of the super-game, (iii) Other team will cooperate unconditionally, and (iv) None of the above. Dialogues were coded for this in each super-game.[35] No team was coded under X5 (iii), believing there were unconditional cooperators among the other teams.

Table 6 reports the evolution of these beliefs over super-games in relationship to the frequency with which teams got up on a cooperative path. Beyond super-game 1 the percentage of teams up on a cooperative path was at least twice as large as the number anticipating conditional cooperators throughout the super-game. The beliefs about the percentage of teams who were strict conditional cooperators drops to single digits in the last three super-games, along with a majority of teams getting up on a cooperative path. This is totally inconsistent with the driving force behind reputation building in Kreps et al.

[Insert Table 6 here]

Conclusion 7: Looking directly at beliefs regarding the likelihood of conditional cooperators in the population shows that far fewer teams' anticipated conditional cooperation compared to the number of teams getting up on a cooperative path. And cooperation increased as the number of teams anticipating defection at some point increased, in direct contradiction to the focal mechanism underlying cooperation in FRPD games in Kreps et al. No team was coded as believing there were unconditional cooperators among the other teams. IV. Discussion

The obvious question from the results reported here is: What drives early cooperation in FRPD games, followed by defection and very gradual (incomplete) unraveling over time? Our data suggests that agents are keenly aware, very early on, that their opponent is likely to cooperate if they do, thereby yielding higher earnings. There is also anticipation, either at the time they start cooperating, or once up on a cooperative path, that their rivals are likely to defect as the end game draws near. So why doesn't this unravel the way the rational agent model

[35] Category X5 used a different set of coders, following similar procedures to those discussed earlier. They had a 70% agreement rate, with disagreement between categories i and ii at 8%, with the remaining disagreements resulting from one coding i or ii with the other coding iv.

predicts with or without "crazy" types? The answer is that agents are boundedly rational, often not responding to defection by defecting earlier in the next stage game, and when responding, typically defecting only one period earlier. Team discussions make it clear that the latter is driven by the failure to consistently account for the fact that other teams are thinking along the same lines as they are. This is strikingly consistent with the level- *k* learning literature, where it is rare indeed to identify any level – 3 types (see, Crawford et al., 2013, for example). This is also supported by the Bereby-Meyer and Roth (2006) experiment demonstrating that principles of reinforcement learning from psychology, which do not rely on agents' cognition, impact cooperation in FRPD games in predictable ways. A complete model to fully account for these characteristics, in conjunction with players' mistakes and initial naiveté (e.g., cooperating in the last stage game) remains to be developed.

A number of interesting questions remain to be explored in FRPD games using the teams technology. First, teams appear to develop the mature pattern of play characteristic of FRPD games faster than individuals and unravel more than for individuals over time. As such, it would be interesting to see how far teams unravel with more experience than is reported here. Will they hit a stationary point or continue to the point of complete unraveling? No doubt this will take bringing experienced subjects back into the lab, or having subjects who are hardy enough to stay alert for a four hour session or longer. Second, given that teams start out cooperating less than individuals, only to cooperate more with a modicum of experience, it would be interesting to try to replicate these results using between-agent discussions, which generates the strongest discontinuity effect reported in the psychology literature.

References

Andreoni, J. and Miller, J. H. 1993. “Rational Cooperation in the Finitely Repeated Prisoner’s Dilemma: Experimental Evidence” *The Economic Journal*, 103 (418), pp. 570-585

Bereby-Meyer, Y. and Roth, A. E. 2006. “The Speed of Learning in Noisy Games: Partial Reinforcement and the Sustainability of Cooperation.” *American Economic Review*, 96 (4), pp. 1029-1042.

Camerer, C. and Weigelt, K. 1988. “Experimental Tests of a Sequential Equilibrium Reputation Model.” *Econometrica*, 56 (1), pp. 1-36

Charness, G. and Sutter, M. 2012. “Groups Make Better Self-Interested Decisions.” *Journal of Economic Perspectives*, 26 (3), pp. 157-176.

Cooper, R., DeJong D., Forsythe, R., and Ross, T. W., 1996. “Cooperation without Reputation: Experimental Evidence from Prisoner’s Dilemma Games.” *Games and Economic Behavior* 12 (2), pp. 187-218.

Cox, C.A., Jones, M.T., Pflum, K.E., Healy, P.J., 2012. “Revealed Reputations in the Finitely-Repeated Prisoners’ Dilemma.” Ohio State University working paper.

Crawford, V. P., Costa-Gomez, M. A. and Iriberri, N., 2013. “Structural Models of Mon-equilibrium Strategic Thinking: Theory, Evidence, and Applications.” *Journal of Economic Literature*, 51 (1), pp. 5-62.

Dál Bo, P., and Fréchette, G., 2011. “The Evolution of Cooperation in Repeated Games: Experimental evidence.” *American Economic Review*, 101 (1), pp. 411-429.

Davis, James, H**.** 1992. “Some Compelling Intuitions About Group Consensus Decisions, Theoretical and Empirical Research, and Interpersonal Aggregation Phenomena: Selected Examples, 1950-1990.” *Organizational Behavior and Human Decision Processes*, 52(1), pp. 3-38.

Fudenberg, D., and Maskin, E. 1986. “The Folk Theorem in Repeated Games with Discounting or with Incomplete Information.” *Econometrica*, 53 (2) pp. 533-554.

Jeheil, Philippe 2005. “Anlogy-based Expectation Equilibrium, ” *Journal of Economic Theory*. 123, pp 81-104.

Jung, Y. J., Kagel, J. H. and Levin, D. 1994. “On the Existence of Predatory Pricing: An Experimental Study of Reputation and Entry Deterrence in the Chain-Store Game.” *RAND Journal of Economics.* 25 (1) pp. 72-93

Kugler, T., Kausel, E. and Kocher, M. 2012. "Are Groups More Rational than Individuals? A review of Interactive Decision Making in Groups," Wiley Interdisciplinary Reviews: Cognitive Science, 3 (4), pp. 471-482.

Kreps, D. M., Milgrom, P., Roberts, J., and Wilson, R., 1982. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." *Journal of Economic Theory*, 27 (2), pp. 245-252.

Lorge, I. and Solomon, H. 1955. "Two Models of Group Behavior in the Solution of Eureka-Type Problems." *Psychometrika*, 20 (2), pp. 139-148.

Neyman, Abraham. 1985. "Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoner's Dilemma Game," *Economics Letters*, 19, pp 227-229.

Reny, Phillip. 1992 "Rationality in Extensive-Form Games." *Journal of Economic Perspectives* 6 (4), pp 103-118.

Reuben, E., Suetens, S., 2012. "Revisiting strategic versus non-strategic cooperation." *Experimental Economics* 15 (1), pp. 24-43.

Selten, R., Stoecker, R., 1986. "End Behavior in Sequences of Finite repeated prisoner's dilemma supergames: a learning theory approach." *Journal of Economics Behavior and Organization* 7 (1), pp. 47-70.

Wildschut, T., Insko, C.A., 2007. "Explanations of Interindividual-Intergroup Discontinuity: A Review of the Evidence." *European Review of Social Psychology* 18 (1), pp. 175-211

Wildschut, T., Pinter, B., Vevea, J.L., Insko, C.A., and Schopler, J., 2003. "Beyond the Group Mind: A Quantitative Review of the Interindividual-Intergroup Discontinuity Effect," *Psychological Bulletin* 129 (5), pp. 698-722.

Windschitl, P. D., Kruger, J., and Simms, E. N. 2003. "The influence of egocentrism and focalism on people's optimism in competitions: when what affects us equally affects me more." *Journal of Personality and Social Psychology*, 85(3), 389-408.

Figure 1

Stage Game Payoffs

(in ECUs)

	A	B
A	105 105	5 175
B	175 5	75 75

Figure 2

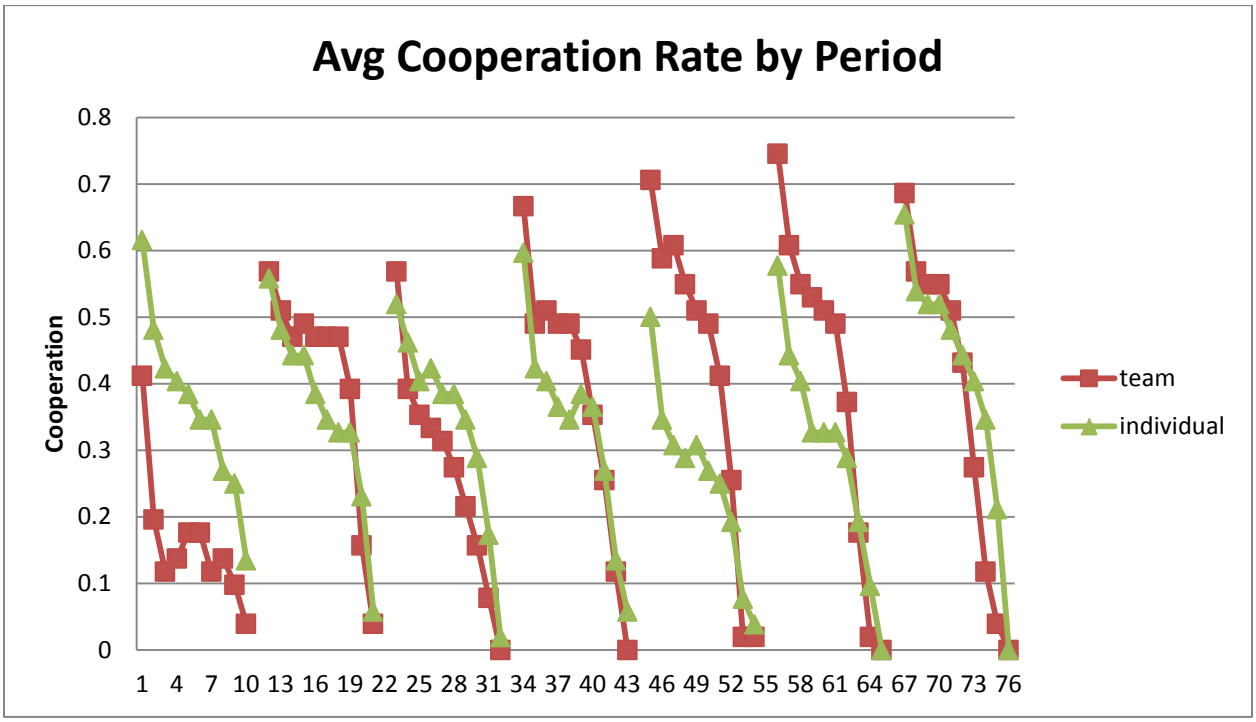

Table 1

Average Stage One Cooperation Rates

Super Game	Individuals	Teams	Difference: Individuals less Teams (z-statistic)
1	0.615	0.412	0.203 (2.07)**
2	0.558	0.569	-0.011 (-0.11)
3	0.519	0.569	-0.050 (-0.51)
4	0.596	0.667	-0.071 (-0.75)
5	0.500	0.706	-0.206 (-2.15)**
6	0.577	0.745	-0.168 (-1.81)**
7	0.654	0.686	-0.032 (-0.347)

** Significantly different from zero, two-tailed test of differences in proportionality.

Table 2
Round Defected in Conditional on Being Up on a Cooperative Path

Super-Game Number	Individuals		Teams	
	Number of Defections[a]	Median (mean)	Number of Defections[a]	Median (mean)
1	14	10 (9.9)	5	9 (9.4)
2	16	10 (9.6)	16	9 (9.1)
3	13	9 (9.0)	9	8 (8.1)
4	9	9 (8.7)	14	9 (8.2)
5	11	9 (8.9)	17	8 (8.0)
6	11	9 (8.8)	16	8 (7.7)
7	15	9 (8.9)	13	7 (7.2)

[a]In cases where both agents defected in same round, both are counted. In cases where one agent defected first, it is counted as a single defection.

Table 3

Change in Round of Defection between Super-Games up on a Cooperative Path

(number of observations)

Earlier Super-game	Were Cooperating		Both Defected		Unilateral Defection		Pooled	
Following Super-game	Individual (20)	Team (35)	Individual (21)	Team (23)	Individual (28)	Team (37)	Individual (69)	Team (95)
Earlier	5.0%	11.4%	38.1%	43.5%	21.4%	10.8%	21.7%	18.9%
Same	35.0%	34.3%	38.1%	21.7%	39.3%	59.5%	37.7%	41.1%
Later	45.0%	31.4%	4.8%	8.7%	25.0%	16.2%	24.8%	20.0%
Same or later	15.0%	22.9%	19.0%	26.1%	14.3%	13.5%	15.9%	20.0%

Table 4

Cooperative Path Patterns

	1st Time up on Cooperative Path		After 1st Time	
	Teams	Individuals	Team	Individuals
P_{CC}	34	36	104	76
P_{DC}	8	1	2	1
P_{DC*}	5	2	3	4

P_{CC} – CC in all of rounds 1-4. P_{DC} – DC (R1), CD (R2), CC rounds 3 and 4.

P_{DC*} – DC (1), CC rounds 2-4

Table 5
Coding Categories: Team Dialogues

Cooperate: Coding conditional on team cooperating (choice of A)

C1. If we cooperate other team might/will cooperate – includes cooperation will result in making more money or necessary to get the other team to cooperate.

C2. What to do if the other team fails to reciprocate cooperation in early plays of the game. Must include reference to defecting at some point in response to the other team's failure to reciprocate.

C3. It's in our best interest to cooperate without discussion of the logic behind cooperating. Essentially C1 above but without discussion of the underlying logic.

C4. Discussion of when to defect in later rounds (including coding the round in which planning to defect).

C5. Partner disagreeing with cooperation – advocating defection.

Defection: Coding conditional on teams defecting (choice of B).

D1. It's the safest choice

D2. Discussion of defection in terms of being a strategic response to the other team's defecting.

D3. Defecting but planning to cooperate if other does so. Often recognize must pay penance as the other team is likely to punish them for having defected. This is only coded for rounds 1-3.

D4. It's in our best interest – defection without any logic behind the doing so.

D5. Recognizing they can't cooperate until the start of a new match, along with the benefits of mutual cooperation. Includes discussion of what to do if the other team fails to reciprocate cooperation in early plays of the game. Analogue to C1 and C2 above.

D6. Partner disagreeing with defection - advocating cooperation.

Additional coding categories irrespective of choices:

X1. I know this game and the way it's supposed to be played; includes having played the same game in a previous experiment or learned about it in a class.

X2. When not cooperating discussing defection in later rounds of a match if and when table to achieve mutual cooperation. Coded just for the first time this occurred.[1]

X3. Partial unraveling - discussion of defecting earlier than in a previous super-game.

X4 Laying out the complete unraveling argument.

X5 Coding for beliefs regarding other team's behavior in *later* plays of the game. Subcategories: (i) Other team is clearly going to defect, or likely to defect, at some point, (ii) Other team *will* cooperate, or is likely to cooperate as long as we cooperate, (iii) Other team will cooperate unconditionally, and (iv) None of the above.

[1] This is the analogue to C4 when the team was not cooperating and was only coded for the first occurrence prior to having assigned C4 to a team. It was done after the initial coding of the data in order to fill an obvious gap in the analysis. It was done by one of the coders.

Table 6

Teams' Beliefs Regarding Other Teams Intentions to Defect and Cooperate in End Stage Games

		Beliefs Regarding Other Teams Actions[a]	
Super-Game	Teams on a Cooperative Path (number)	Percent Defect	Percent Cooperate
1	7.8% (4)	25.0%	25.0%
2	47.1% (24)	40.9%	18.2%
3	31.4% (16)	43.7%	12.5%
4	49.0% (25)	60.0%	12.0%
5	52.9% (27)	57.7%	3.8%
6	51.0% (26)	61.5%	3.8%
7	54.9% (28)	50.0%	7.1%

[a]Percentages are based on agreement between coders.

Appendix – Results from Round 1 Cooperation Rate Probits

The dependent variable takes the value 1 if an agent cooperated in round one of each super game. For explanatory variables we looked to the key elements characterizing first round play in Dál Bo and Fréchette (2012). Comparable forces are at work here: The initial cooperation rate dummy is equal to 1 if the agent cooperated in round 1 of super-game 1, included to capture agents' inherent tendency to cooperate or not. A team dummy is introduced to account for differences in cooperation rates between teams and individuals (value of 1 for teams). Super game is a linear time trend across super games. The immediate past experience dummy takes on a value of 1 if the agent they were paired with in previous super game cooperated. Two specifications are reported, with and without interaction effects between the teams dummy and the other explanatory variables.

Absent any interaction variables for teams with the other right hand side variables, both the initial cooperation rate and immediate past experience dummies are significant at the 1% level. The teams dummy is positive and significant at the 10% level, and the time trend variable is positive and significance at the 5% level.

A specification including interaction terms between the teams dummy and each of the other right hand side variables showed none to be statistically significant on their own, at anything approaching conventional levels, with essentially no overall effect as well (χ^2 (3) = 0.55).

Probit Regression: Cooperation Rates Across Super-games

(Standard errors corrected for clustering at the subject level)

Constant	-0.770 (0.257[a]
Initial cooperation rate dummy	0.772 (0.213)[a]
Team dummy	0.399 (0.214)[c]
Super game	0.059 (0.029)[b]
Immediate past experience dummy	0.374 (0.094)[a]
Pseudo Log-Likelihood	-376.9
Pseudo R^2	0.097

[a] Significantly different form 0 at the 1% level

[b] Significantly different form 0 at the 5% level

[c] Significantly different form 0 at the 10% level