Cooperation through Communication:
Cheap Talk in a Finitely Repeated Prisoners’ Dilemma Game*

John H. Kagel
Department of Economics
Ohio State University
5/29/2016

Abstract

For both two person teams and individuals cheap talk in finitely repeated prisoner dilemma games results in a huge increase in stage-one cooperation rates compared to its absence: 70.6 and 49.4 percentage point increases for individuals and teams, respectively. Between opponent chats focus on the increased earnings cooperation can achieve, with minimal discussion of punishment for defection. However, to restore cooperation after an early stage-game defection typically requires compensating the aggrieved player. We revisit the “discontinuity effect” reported in the psychology literature that teams are less cooperative than individuals.

Key words: finitely repeated prisoners’ dilemma game, cheap talk, team versus individual play, discontinuity effect.
JEL classification: D03, C92, C73

*Research support from national Science Foundation grant SES-1226460 is gratefully acknowledged. We thank seminar participants at the University of Pittsburgh and the University of Michigan for their comments which have improved the paper substantially. We alone are responsible for any errors or omissions.
Results are reported for an experiment investigating behavior in a finitely repeated prisoners’ dilemma game (FRPD) where opponents can talk to each other between stage-games, but cannot make binding agreements (cheap talk). Communication of this sort is a central element underlying cooperation in a variety of repeated games outside the laboratory: It is a common element to collusive arrangements within cartels (e.g., Genosove and Mullin, 2001), collaboration with colleagues, and coordination games, to name but a few. Laboratory studies of the role of communication between agents in conflict situations has been drawing some attention lately (Fonseca and Norman, 2012; Cooper and Kühn, 2014; Arechar et al., 2016). While with common knowledge of rationality, FRPD games should unravel to eliminate all cooperation, as is well known this is not the case for games of any length (e.g., Selten and Stoecker, 1986; Andreoni and Miller, 1993). This can be justified by a breakdown in common knowledge of rationality as a result of various types of “craziness” (Reny, 1992, Kreps et al., 1982).

We study the effect of cheap talk for both teams and individuals as (i) many economic decisions are made in teams and (ii) through analyzing within team discussions, it provides an opportunity to understand what motivates behavior. In the latter context there is an added bonus in that we can compare the content of cheap talk between opponents for individuals and teams to determine the extent to which similar behavioral processes are at work. In using teams and individuals we can also investigate whether teams are consistently less cooperative than individuals as reported in the social psychology literature on repeated play prisoner dilemma games (referred to as the ”discontinuity effect”; see Wildschut et al., 2003 and Wildschut and Insko, 2007, for surveys of this literature). The “discontinuity effect” has been used to argue that teams are inherently more aggressive and less trusting (and trustworthy) than individuals both in the psychology and economics literature (Charness and Sutter, 2012; Kugler et al., 2012).

Focusing on cooperation rates in the first stage-game of a sequence of finitely repeated prisoner dilemma (FRPD) super-games, we find large and consistent increases in cooperation compared to the absence of cheap talk, for both individuals and teams: Average cooperation rates of 62.2% absent communication to 92.9% with communication for teams and average cooperation rates of 57.4% absent communication to 97.9% with for individuals, with most stage-one defections occurring in the first super-game. For both teams and individuals, between agent discussions prior to each stage-game focus on the increased earnings from cooperating, with fairness coming in second, and punishment if their opponent does not cooperate coming in a
distant third. Restorations of cooperation following breakdowns in early stage-game cooperation are almost always associated with explicit compensation for the agent earning the sucker payoff. For both teams and individuals, unilateral defections over the last several stage-games are most often met with no comment or mild upset, with real upset or sarcasm occurring half as often.

Average stage-one cooperation rates across super-games with communication show a small (5 percentage point), higher cooperation rate for individuals consistent with the discontinuity effect and that individuals are more cooperative than teams. In contrast, there is a small, 4.8 percentage point, difference in the opposite direction with no communication. However, both with and without communication, stage-one cooperation rates in the initial super-game are higher for individuals, suggesting that the discontinuity effect may be limited to early super-games.¹ At the same time, both with and without communication, individuals are much more likely to cooperate in the last stage-game than teams, consistent with the notion that teams are more “rational” as that term is commonly used in economics.

The remainder of the paper is organized as follows: Section I outlines our experimental design and procedures. Section II reports the experimental results. Section III briefly summarizes the main results.

II. Experimental Design and Procedures:

Procedures are first described for the cheap talk sessions. They were essentially the same for games without cheap talk, with the differences described briefly at the end of this section.

Unrestricted communication between opponents was permitted prior to each stage-game via a continuous chat box, with a separate chat box for within team discussions where applicable. In the teams treatment either member of each team could communicate with their common opponent, with each teammate privy to all between team discussions.² Subjects were instructed to not use profanity or to reveal who they were, which essentially never happened in spite of our having no formal mechanism to prevent violations. We chose unrestricted communication as it would seem to correspond most closely to the nature of communication outside the lab in environments with repeated interactions. What data there is available suggest that unrestricted

---

¹ As reported on below, in repeated prisoner dilemma games, psychologists typically employ a single super-game. The effect of communication between opponents reported in the psychology literature is briefly reviewed below.
² Early on there was occasional confusion on the part of teammates so that within team discussions were recorded in the between team chat box. This was quickly corrected when it occurred.
communication results in higher cooperation rates than restricted communication (e.g., choosing
between pre-programmed messages; Cooper and Kühn, 2014).

Subjects played a ten 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 total payoff.

[Insert Figure 1 here]

In the team treatment, subjects were randomly matched with a partner at the beginning of
an experimental session, with partners remaining the same throughout the session. Teams played
against teams, and individuals played against individuals. In what follows we will use the term
agent or player to refer to either a two person team or an individual. Following each FRPD
game, agents were randomly and anonymously re-matched under the restriction that no two
agents would be re-matched in consecutive super-games. All teams played in seven FRPD
super-games which were about all we could squeeze in a two hour session. All individuals
played in ten FRPD games. Agents in both treatments were told they would play between 7 and
10 super-games.

For teams, each stage-game began with a brief period of within team discussions,
followed by a period for between team discussions. This was followed by a brief period for
within team chat during which teams decided whether to defect or cooperate. All
communication, both within and across teams, was done anonymously through computers, with
subjects sitting at their own computer terminals. For individuals, each stage-game began with
cheap talk between opponents, after which decisions were made. Here too all communication
was done anonymously through computers. The time available for each of these stages of play is
detailed in the instructions, along with the default options if teams failed to reach agreement
(which was rare). 3

Following the end of each stage-game agents had up to 30 seconds to view their own
results before moving on to the next stage-game, where again teammates first had an opportunity
to talk to each other, followed by between team discussions, and ending with within team

3 See http://www.econ.ohio-state.edu/kagel/cooperation through communication.
discussions and deciding what action to take. 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 B in each stage-game, and were told they would be “paired with the same other team (individual) for a set of 10 repeated choices.”

With cheap talk there were three team sessions with between 8 and 10 teams in each for a total of 28 teams and three individual subject sessions with between 10 and 12 subjects for a total of 34 individuals. Sessions without cheap talk were essentially the same except for doing away with the cheap talk between opponents. There were 5 sessions with teams and 5 with individuals absent cheap talk with between 8 and 12 individuals/teams in each session. These sessions were conducted approximately one and a half years before the cheap talk sessions. Results from these sessions are reported in detail in Kagel and McGee (2016; hereafter KM) and will be used primarily to compare outcomes with the cheap talk sessions.

II. Theoretical predictions

Assuming common knowledge of rationality in FRPD games, the standard theoretical prediction is that there will be no cooperation between agents using a backward induction argument. As is also well known this virtually never happens for super-games of any length which are typically characterized by some early stage-game cooperation only to trial off and end as the last stage-game approaches. Given these initial tendencies to cooperate in early stage-games, following Farell and Rabin (1996) communication should aid in achieving cooperation in early stage-games via self-signaling (when the sender prefers the receiver to play the message sent if the sender truly intends to play the signaled action). Further, what communication can do in FRPD games is to help agents overcome the fear of getting the sucker payoff, at least in early stage-games. Further, when agents fail to live up to their promises early on, communication prior to the next stage-game might also help to overcome the initial breach of confidence, say by coordinating on “getting even” or by balling out their opponent. As such we would predict increased cooperation in early stage-games with communication than without, which would be sustained for a longer number of stage-games.

4 Procedures were the same for individuals absent the need for within team discussions.
5 One these team sessions used a student assistant to ensure an even number of teams. The assistant informed his teammate that he was part of the experimental team and would agree to whatever his partner did. Data for this team is dropped except as needed to complete play when paired with another team.
The psychology literature on repeated PD games shows that teams are less cooperative than individuals, an outcome referred to as the “discontinuity effect.” These social psychology experiments are conducted under very different procedures than those usually employed in economics. In the social psychology experiments subjects are typically told that they will be paired with the same opponent for between xx and yy plays of the game (e.g., between 4 and 10 plays), with the resulting super-game ending somewhere before the announced upper bound is reached. Further, subjects typically play in a single super-game. In the social psychology literature teams will consist of two, or often more than two, subjects with different rules for determining team choices (typically a consensus choice). Financial incentives, typical to those used in economic experiments are commonly employed (see Wildschut and Insko, 2007, and Wildschut et al., 2003, for surveys of the literature). Experiments are conducted with and without communication between opponents, with the later usually involving face-to-face communication between a single representative from each team. Further, cooperation rates are commonly averaged across all of the stage-games, whereas economists focus on cooperation rates in the initial stage-game.

Within this literature the most common explanation for why teams are less cooperative in PD games is that they are more fearful and less trusting than individuals, preferring the “safe” choice (defection). KM support this argument in that 91.7% (22/24) of the teams failing to cooperate in the first stage-game of the first super-game were coded as doing so on the basis of defection being the safe and/or needed to avoid the sucker payoff. Wildschut et al. (2003) in reviewing the literature note that, if anything, communication between opponents will, magnify the difference in cooperation rates between teams and individuals, with the most dramatic differences obtained with unrestricted communication.6

III.1 Cheap Talk Results:

Figure 2 compares cooperation rates within and across super-games for individuals with and without communication. The most striking result is the huge increase in early stage-game cooperation with communication. Further, outside of the first super-game, stage one cooperation

---

6 “Consistent with our initial expectation regarding the role of communication, studies without communication were associated with the smallest (discontinuity) effect sizes and studies with unconstrained communication were associated with the largest effect sizes” (Wildschut et al., 2003, p. 709, parentheses added).
rates are 100% with communication, which almost always persists through the first several stage-games, only to trial off as the end stage draws near. Noticeably, with communication end game cooperation rates start out high in the first super-game and never got below 26% in subsequent super-games. In contrast, absent communication, the maximum frequency of end game was 13.5% in the first super-game, and 0% in the last two super-games.

Figure 3 reports the corresponding results for teams. Here too there is a huge increase in stage one cooperation rates with communication than without. However, except for super-game six, average stage-one cooperation rates are always a bit lower than 100%. Further, after the first super-game, end game cooperation rates are close to, or equal to, zero with communication, although here too the rates are noticeably higher than absent communication.

Table 1 reports average stage one cooperation rates for each super-game with communication along with z-statistics for differences between the two treatments. The focus is on stage-one cooperation rates, as cooperation in later rounds is dependent on what happens in the first stage-game, as defection sets off complicated interdependencies that are difficult to account for. Further, once two or more stage-games have been completed in which one of the agents has defected, in the overwhelming number of cases, both agents defect for the remainder of the super-game.

With communication, average stage-one cooperation rates for teams in super-game one are lower than for individuals, but the difference is relatively small and not statistically significant (75.0% for teams versus 83.5% for individuals; p > 0.10, Fisher’s exact test).\(^7\) This is smaller than the difference absent communication (41.2% for teams versus 61.5% for individuals; p < 0.05). Looking beyond the first super-game, the fact that with communication stage-one cooperation rates are 100% in all subsequent super-games for individuals, while less than 100% for teams, treating stage-one behavior across super-games as independent, there is a small, but statistically significant (p < 0.01), difference in cooperation rates between teams and

---

\(^7\) Unless noted otherwise, all proportion terts use Fisher’s exact text.
individuals (92.9% for teams versus 97.9% for individuals).\textsuperscript{8} In contrast, averaging across super-games without communication, teams are more cooperative (62.2% versus 57.4%, p < 0.10).\textsuperscript{9}

\textit{Conclusion 1:} For both teams and individuals, the primary impact of introducing unrestricted cheap talk into FRPD games is a large and significant increase in stage-one cooperation rates. In both cases, teams are less cooperative in the first super-game, but the difference is smaller with communication than without, and is statistically significant (insignificant) without (with) communication.

Fully rational, own income maximizing agents should never cooperate in the last stage-game as it is a dominant strategy to defect. In this respect, with communication teams are much more rational than individuals as 32.1% cooperated in the last stage at least once compared to 79.4% for individuals ($Z = 3.75, p < 0.01$).\textsuperscript{10} The number of repeat offenders is also substantially higher for individuals with 67.6% of individuals cooperating in the last stage-game more than once versus 11.1% for teams. These high end game cooperation rates contrast with the substantially lower rates without communication: 26.9% for individuals verses 9.8% of the teams.\textsuperscript{11}

Cooperation in the last round is sometimes treated as evidence for either pure altruists or conditional cooperators (Andreoni and Miller, 1993; Cox et al., 2015). However, the data suggests otherwise as with communication all the teams cooperating in at least once end game, either unilaterally defected, or defected simultaneously with their opponent in another super-game, and all but four individuals did so.\textsuperscript{12} This rules out pure altruism, or conditional cooperators, as being responsible for the much higher end game cooperation rates for individuals. Learning across super-games, and reacting to being defected on in previous stage-games, has an obvious effect on whether or not there is cooperation in the last stage-game. A

\textsuperscript{8} Based on an exact binomial test statistic. The 100% cooperation rate for individuals after the first super-game precludes more complicated probits such as those reported in Däl Bo and Fréchette (2011) and KM that show that initial tendencies to cooperate, and whether their opponent cooperated or not in stage-one of the previous super-game, affect the probability of cooperating in stage-one of the present super-game. That is to say, the assumption that outcomes are independent underlying the binomial test statistic are not exactly satisfied.

\textsuperscript{9} See the probit reported in the Appendix to KM.

\textsuperscript{10} For both treatments end game cooperation rates were highest in the first super-game.

\textsuperscript{11} Absent cheap talk no team cooperated more than once in the last stage-game, with xx% of individuals doing so.

\textsuperscript{12} Without communication all teams and all individuals initiated either initiated defection or did so simultaneously in a subsequent super-game.
probit confirms this as, with communication, end stage cooperation is half as likely to be repeated after it was met with defection, as opposed to cooperation, in the previous super-game.

**Conclusion 2:** With communication there is significantly less cooperation in the last stage-game for teams than for individuals. The same is true absent communication, with these rates substantially lower than the corresponding rates with communication. There is little evidence that end game cooperation is driven by altruism or committed cooperators, as for all but a handful of players, they either unilaterally defect, or defect simultaneously, on or before the end game in or more super-games.

Cooperation unravels more and faster for teams than for individuals conditional on being up on a cooperative path at the start of a super-game. To measure this we identify the stage-game in which an agent first defects, conditional on being on a cooperative path at the start of a super-game. The latter is defined as sustained cooperation over rounds 1–4, typically with both agents cooperating in all rounds.\(^\text{13}\) Table 2 reports the median, and average, stage-game in which these first defections occurred across super-games both with and without communication. In both cases the median in the first super-game is the last stage-game (10) for individuals versus the next-to-last stage-game (9) for teams.\(^\text{14}\) There is slow, and far from complete, unraveling across super-games in both cases, with the median for teams always one step ahead of individuals until the next to last super-game, where it is two steps ahead (stage-game 7 versus 9). Comparing counts across teams and individuals, the difference in medians in the last super-game are significant in both cases (p < 0.01 using a Mann-Whitney test statistic).

**Conclusion 3:** The pattern of unraveling over time is essentially the same both with and without communication. Cooperation unravels faster for teams than for individuals.

\[\text{[Insert Table 2 and 3 here]}\]

With communication, the most common pattern by far for getting up on a cooperative path was when there was mutual cooperation over the first four stage-games. However, within the first four stage-games there were a number of unorthodox patterns under which agents were able to sustain cooperation for a large number of stage-games until the defections reported in

---

\(^{13}\) There are exceptions to this criterion discussed below.

\(^{14}\) A unilateral defection counts as one observation in Table 3, with mutual defection counting as two observations.
Table 2. These unorthodox patterns were most common in the first super-game and more common for teams than individuals (see Table 3). One of these unorthodox patterns had agents alternate between defection and cooperation within the first three stage-games, only to restore cooperation for a number of stage-games. Restoring cooperation typically required “evening the score”, either explicitly following the initial defection, or explicitly referred to after the cooperating player retaliated: Of 19 super-games with unilateral defection within the first 3 stage-games, 13 “evened the score”, trading defection and cooperation, with 10 of these restoring cooperation. Only 1 out of the 6 super-games that failed to even the score returned to cooperation.16

III.2 Content Analysis of within Team and Between Agent Discussions

The results reported show a huge increase in initial cooperation rates after allowing for communication between players. This section looks at these between opponent discussions to better understand the basis for the increase in cooperation rates, as well as the within team discussions to better understand the factors motivating team play. Further, to the extent that between opponents discussions are similar for teams and individuals, this provides some confidence that similar factors motivate individual subject play, which is not directly observable.

Two students were used to code within and between agent discussions according to pre-specified categories of general interest (e.g., agents’ responses to end game defection), and after I read a sample of the chats with these categories in mind. Coders could assign multiple codes to the same stage-game. The coders independently coded a common session, after which we met jointly to refine our common understanding of what the categories were designed to capture. Coders then independently coded the rest of the sessions, after which there was a final meeting to reconcile obvious discrepancies.

Table 4 reports the between player dialogues for individuals and teams in the first stage-game for super-games 1, 3 and 6. The goal here was to identify the principle arguments for cooperating – fairness, way to make the most money, and threats should their opponent fail to cooperate. Between opponents chat had declined substantially by super-game 6, with brief discussions largely limited to agreeing to cooperate, so that most of the codes were assigned in

15 Similar patterns are reported absent communication (see KM, 2016).
16 One example of explicitly agreeing to settle the score: From the team defected on “Seriously if you choose A (cooperate) we all win”. From the team that unilaterally defected “we'll go A now and you guys take B (defect) so its even”.

Page 10
super-game 1, and 3 to a lesser extent. The numbers report the frequency with which at least one coder coded the category in question.\textsuperscript{17} Snippets of discussions for each category are reported at the bottom of the table.

The most frequently coded category consisted of cooperating to get the most money for both players, accounting for a little over 50\% for both teams and individuals. Fairness followed a close second at 30-40\%, with threatening their opponent, should they not cooperate, coming in a distant third at 8\%. While the high frequency of appealing to the mutual benefits of cooperation is not surprising given previous results reported, the near non-existence of threats for cheating is much less common.\textsuperscript{18}

[Insert Table 4 here]

We also coded within team chats \textit{after} the between team discussions, when teams were deciding whether to cooperate or not. The extant psychology literature on the discontinuity effect argues that major concerns underlying the failure of cooperation for teams is “distrust” and/or “safety” considerations, which we coded for, along with any discussions of retaliation if the other team failed to cooperate. This too was coded for the first stage-game of super-games 1, 3 and 6, with multiple codes permitted for a given stage-game. Table 5 reports these frequencies. Distrust regarding their opponent’s intention to cooperate is the most frequently coded category by far, at just shy of 74\%, with most of these discussions occurring in super-game 1. However, teams typically did not act on these concerns, as evidenced by the high stage 1 cooperation rates in the first super-game (75\%). Here too, discussions of retaliating should the other team not cooperate were minimal, perhaps because of an implicit understanding that cheating would result in retaliation, which almost always happened.

[Insert Table 5 here]

\textit{Conclusion 4:} Between player discussions for both teams and individuals focused on cooperating to make the most money, along with fairness considerations. Threats if their opponent failed to cooperate came in a distant third. Within team discussions prior to choice were focused on

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
Category & Frequency & Notes \\
\hline
Making the most money & Team: 90\%, Individual: 85\% & Agreement rates for the most frequently coded category, “making the most money”, were 90\% for teams and 85\% for individuals. \\
\hline
Fairness & Team: 30-40\%, Individual: 30-40\% & Agreement rates for coders summed over the three categories was 75\% for teams and 68\% for individuals. \\
\hline
Threats & Team: 8\%, Individual: 8\% & For example, Cooper and Kühn (2014) in their treatment with unconstrained communication in a coordination game prior to individuals choosing simultaneously, report appeals to mutual benefits just shy of 30\%, with threats of punishment for defecting to an outcome where one player would be better off, at just under 20\%. \\
\hline
\end{tabular}
\caption{Frequency of coded categories. }
\end{table}
“distrust” as to whether their opponent follow through on agreeing to cooperate. Most of the latter discussions occurred in stage-one of super-game one and, as the data shows, did not result in wholesale defections.

Table 6 reports messages sent to their opponent following unilateral defection in stage-games 7-9. Making no comment or expressing mild upset following defection is coded most often for both teams and individuals (55.3% and 60.0%, respectively). Sarcasm or real upset occurring a bit more than half as often, and did not materially slow down the slow unraveling of cooperation over time compared to the absence of any opportunity to communicate disapproval (recall Table 2). The latter is somewhat surprising, since in a voluntary contribution game Masclet et al. (2003) report that giving agents the opportunity to express disapproval of others’ decisions consistently increased contribution levels, and early on increased them to levels observed with monetary punishment. Perhaps the logical inevitability of defection as the end game draws near drives the ineffectiveness of verbal disapproval here. Although one would think the same logic holds for a VCM game with the stranger matching protocol reported in Mascelt et al.

[Insert Table 6 here]

Teams were coded for the first super-game in which they discussed defecting while cooperating in early stage-games. These first discussions occurred primarily in super-games 1 and 2 (67.8% and 25.0%, respectively). In both cases most of these discussions occurred on or before the fourth stage-game (68.4% and 85.7% for super-games 1 and 2, respectively), and with one exception, the proposed round for defection was either the last, or next to last stage-game, with 83.3% discussing defection in the last stage-game. However, about half of the teams discussing defection in the last stage-game, actually defected in the next to last stage-games, indicating some limited backward induction as a result of further within team discussions. A

---

19 For super-games 1, 3, 6. Frequency with which the category was coded by one of the two coders. The agreement rate between coders for teams was 96% and 83% for individuals. There was no opportunity for sending messages after the last stage-game.
20 The two coders’ achieved 100% agreement on the super-game in which defection was first discussed, as well as the other characteristics discussed here. Recall, there were 28 teams.
21 The one exception proposed defection in the second stage-game after being defected on while cooperating in the first stage-game.
22 Note, however, that one these teams planning to defect in the last stage-game actually cooperated, and another moved from planning to defect in the next-to-last stage-game, the last stage-game.
few of the teams proposing defection in super-games one and two did so just prior to, or during, the stage-game in which they defected (stage-games 8-10), 15.8% and 14.3% respectively.  

**Conclusion 5:** Most teams first discussed defection well before the end of super-game 1, typically planning to defect in the last stage-game. As such, most teams acted as if they were well aware of the dominant strategy for the last stage-game (although none discussed it formally), but engaged in limited or no backward induction, accounting for their opponents thinking along the same lines.  

IV. Discussion  

This paper investigates the effect of between “opponent” discussions prior to choice in a finitely repeated prisoner dilemma game. For both teams and individuals, the opportunity to engage in unrestricted pre-play communication increased stage-one cooperation rates substantially compared to no communication: from 62.2% to 92.9% for teams and from 57.4% to 97.9% for individuals. Unrestricted communication of the sort employed here typically results in substantially higher cooperation rates than the same games with more limited communication.23 Between opponent discussions underlying the extremely high cooperation rates reported here focus on the increased earning opportunities and fairness from cooperation, with very limited discussions of retaliation for failure to cooperate. However, the latter typically occurred following unilateral defection, suggesting that retaliation was, at least implicitly, a common understanding among players. Unilateral defection in later stage-games (after agreeing to cooperate), were most often met with no comment or expressing mild upset with the opposing party. Harsher comments following these defections, which occurred 32-38% of the time, had little if any effect on slowing down unraveling over time compared to games with no ability to communicate, indicating that these comments had little if any effect on cooperation.  

Teams were much more likely to defect on or before the last stage-game than individuals, both with and without communication, suggesting better understanding of the dominant strategy to do so. Teams also ended joint cooperation earlier than individuals. Within team discussions in the communication treatment show that while cooperating in the first two super-games, most teams were planning to defect on, or just before, the last stage-game. This is consistent with one of the main ideas underlying the Kreps et al. model – that players are able to see through to the  

---

23 See Arechar, et al. (2016) and Wildschut et al. (2003) for prisoner dilemma games, and the between treatment conditions reported in Cooper and Kuhn (2014).
last stage-games, planning to defect at some point, while cooperating early on. While, unlike that model, early stage-game cooperation is developed through a trial and error learning process.24

With communication, teams are consistently less cooperative than individuals across super-games, but the difference is quite small. Like the no communication treatment, the largest difference between teams and individuals occurs in the first super-game (10.3% with cheap talk versus 20.3% without). However, in the no communication treatment, teams are as, or more, cooperative than individuals in later super-games.

The most striking result reported here is the sharp increase in stage-one cooperation rates with communication compared to without. In anti-trust legislation it is against the law to discuss cooperation between competitors, regardless of whether it results in a collusive outcome. Although we recognize that there is no direct link between the experiment reported here and anti-trust issues outside the lab, to the extent that the prisoners’ dilemma serves as a metaphor for competition between the few, the sharp increase in cooperation rates reported here with cheap talk serves to support the prohibition on rivals even discussing cooperation.

Our data also has something to say with respect to the “discontinuity effect” reported in the social psychology literature – the tendency for teams to be less cooperative in repeated prisoner dilemma games. Contrary to what was expected based on the extant social psychology literature, that communication would result in larger differences in the discontinuity effect, average stage-one cooperation across super-games for individuals without communication is actually lower than for teams, and minimal differences in the “right” direction with communication. Further, in the first super-game where the differences between individual and team cooperation rates are largest, with teams cooperating less in both cases, the difference is smaller with communication than without. This suggests that the discontinuity will be concentrated in the initial super-game for repeated PD games, but may well be weaker with unrestricted communication than without.25

It will be interesting to extend the results reported here to indefinitely repeated super-games of the sort reported in Dál Bo and Fréchette (2011) and Fudenberg et al. (2012).

24 See KM for documentation of this effect.
25 This is conditional on stage-game payoffs supporting cooperation in the repeated game. This is an open question since, as noted earlier, the psychology literature typically employs a single super-game.
References


Figure 1
Stage Game Payoffs
(in ECU)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>105</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>105</td>
<td>175</td>
</tr>
<tr>
<td>B</td>
<td>175</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>75</td>
</tr>
</tbody>
</table>

Figure 2. Individuals With and Without Communication.
Figure 3. Teams With and Without Communication
### Table 1

**Average State-One Cooperation Rates**

<table>
<thead>
<tr>
<th>Super-Game</th>
<th>Individuals</th>
<th>Teams</th>
<th>Diff: Indvds-Teams (test-statistic)²</th>
<th>Individuals</th>
<th>Teams</th>
<th>Diff: Indvds-Teams (test-statistic)²</th>
</tr>
</thead>
<tbody>
<tr>
<td>With Communication</td>
<td></td>
<td></td>
<td></td>
<td>Without Communication</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.850</td>
<td>.750</td>
<td>.130 (0.35)</td>
<td>.615</td>
<td>.412</td>
<td>.203 (&lt; 0.05)</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>.964</td>
<td>.036 (0.45)</td>
<td>.558</td>
<td>.569</td>
<td>-.011 (&gt; 0.99)</td>
</tr>
<tr>
<td>3</td>
<td>1.00</td>
<td>.964</td>
<td>.036 (0.45)</td>
<td>.519</td>
<td>.569</td>
<td>-.050 (.69)</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>.929</td>
<td>.071 (0.20)</td>
<td>.596</td>
<td>.667</td>
<td>-.071 (.56)</td>
</tr>
<tr>
<td>5</td>
<td>1.00</td>
<td>.964</td>
<td>.036 (0.45)</td>
<td>.500</td>
<td>.706</td>
<td>-.206 (&lt; 0.05)</td>
</tr>
<tr>
<td>6</td>
<td>1.00</td>
<td>1.00</td>
<td>0.00 (&gt; 0.99)</td>
<td>.577</td>
<td>.754</td>
<td>-.168 (&lt; 0.10)</td>
</tr>
<tr>
<td>7</td>
<td>1.00</td>
<td>.929</td>
<td>.071 (0.20)</td>
<td>.654</td>
<td>.686</td>
<td>-.032 (.83)</td>
</tr>
</tbody>
</table>

²Fisher’s exact test, 2-tailed test statistic.

### Table 2

**Round Defected in Conditional on Being Up on a Cooperative Path:**

<table>
<thead>
<tr>
<th>Super-Game Number</th>
<th>With Communication</th>
<th>Without Communication</th>
<th>With Communication</th>
<th>Without Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individuals</td>
<td>Teams</td>
<td>With</td>
<td>Individuals</td>
</tr>
<tr>
<td></td>
<td>(median) (mean)</td>
<td>(median) (mean)</td>
<td>(median) (mean)</td>
<td>(median) (mean)</td>
</tr>
<tr>
<td>1</td>
<td>10 (9.8)</td>
<td>10 (9.0)</td>
<td>9 (9.3)</td>
<td>9 (9.4)</td>
</tr>
<tr>
<td>2</td>
<td>10 (9.8)</td>
<td>10 (9.3)</td>
<td>9 (9.0)</td>
<td>9 (8.9)</td>
</tr>
<tr>
<td>3</td>
<td>10 (9.7)</td>
<td>9 (9.0)</td>
<td>9 (8.4)</td>
<td>8 (7.9)</td>
</tr>
<tr>
<td>4</td>
<td>9 (9.4)</td>
<td>9 (8.7)</td>
<td>8 (8.3)</td>
<td>9 (8.2)</td>
</tr>
<tr>
<td>5</td>
<td>9 (9.3)</td>
<td>9 (8.7)</td>
<td>8 (7.8)</td>
<td>8 (8.1)</td>
</tr>
<tr>
<td>6</td>
<td>9 (9.2)</td>
<td>9 (8.8)</td>
<td>7 (7.5)</td>
<td>8 (7.8)</td>
</tr>
<tr>
<td>7</td>
<td>9 (9.1)</td>
<td>9 (8.9)</td>
<td>7 (7.1)</td>
<td>7 (7.0)</td>
</tr>
</tbody>
</table>

²In cases where both agents defected in same round, both are counted. In cases where one agent defected first, counted as a single defection.
Table 3
Frequency of Different Cooperative Path Patterns

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Super-Game 1</th>
<th></th>
<th>Super-Games 2-7</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Teams</td>
<td>Individuals</td>
<td>Team</td>
<td>Individuals</td>
</tr>
<tr>
<td><strong>P_{CC}</strong></td>
<td>10</td>
<td>24</td>
<td>150</td>
<td>200</td>
</tr>
<tr>
<td>Other</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>No Coop Path</td>
<td>10</td>
<td>4</td>
<td>12</td>
<td>4</td>
</tr>
</tbody>
</table>

P_{CC} – both players cooperate for the first 4 stage games. Other – see text.

Table 4
Between Player Messages Prior to Stage-One Choices*
(percent distribution across categories)

<table>
<thead>
<tr>
<th>Coding Category</th>
<th>Teams</th>
<th>Individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fairness</td>
<td>36.1%</td>
<td>40.0%</td>
</tr>
<tr>
<td>Most money</td>
<td>55.6%</td>
<td>52.0%</td>
</tr>
<tr>
<td>Threats</td>
<td>8.3%</td>
<td>8.0%</td>
</tr>
</tbody>
</table>

*Coded for super-games 1, 3 and 6.

Chat examples:

**Fairness:**
“I say we all just be fair about it and choose A (cooperate)”
“how do we choose?” “A is more mutual”

**Most Money:**
“yo other team if you trust us we can both choose a (cooperate) and make some hashtag cash”
“listen here. A (cooperate) makes the most money” “makes the most money to pick A” “we wont (sic) the highest earnings”

**Threats:**
“ok guys here is the thing you can say you dont trust us and go with B but you'll lost 30 pints for each round after that”

**Most Money and Fairness:**
“I say we all just be fair about it and choose A. That way everyone makes good money” “lets take the experiment people and both choose a”
Table 5
Within Team Discussions Preceding Stage-One Choices
(percent distribution across categories)

<table>
<thead>
<tr>
<th>Coding Category</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distrust</td>
<td>73.7%</td>
</tr>
<tr>
<td>Safety</td>
<td>15.8%</td>
</tr>
<tr>
<td>Retaliate</td>
<td>10.5%</td>
</tr>
</tbody>
</table>

Chat Examples:

Distrust:
“We going with a (cooperate)?” “I guess so” “hopefully the other team sticks to their word” “That does concern me”

Both Distrust and Safety:
“so the question is do we trust them picking a (cooperate)? If they choose b (defect) and we choose a we’re screwed” “true I don’t know should we?”

Retaliate:
“we decided A (cooperate) right?” “yea” “cool” “if they choose a, too …if not, we probably choose b (defect) from next round”
<table>
<thead>
<tr>
<th>Coding</th>
<th>Teams</th>
<th>Individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>No comment or expressing mild upset</td>
<td>55.3%</td>
<td>60.0%</td>
</tr>
<tr>
<td>Expecting it to happen</td>
<td>6.4%</td>
<td>8.0%</td>
</tr>
<tr>
<td>Expressing real upset or sarcasm</td>
<td>38.3%</td>
<td>32.0%</td>
</tr>
</tbody>
</table>

*Chat Examples:*

*No comment or expressing mild upset:*
  - “ouch”
  - “I thought you were trusting!?”
  - “wow that was interesting”

*Expecting it to happen:*
  - “I saw that coming”
  - “I knew it”

*Expressing real upset or sarcasm:*
  - “ok be a douche”
  - “wow was that worth the 75 ur gonna get the rest of the way now”
  - “well done” “we know your hand must have slipped”