Southern Economic Journal 2016, 82(4), 1167–1185
DOI: 10.1002/soej.12109

Communication and Cooperation: A Methodological Study

David J. Cooper* and Kai-Uwe Kühn†

We study the effect of communication in an experimental game where cooperation is consistent with equilibrium play. We examine two methodological questions which affect many studies of communication in games, particularly those studying the relationship between communication and cooperation in supergames. We show that making it easier for subjects to reach an agreement (by allowing more rounds of communication) does not increase cooperation with a limited message space treatment. Thus, limited message space treatments are not a good substitute for the use of chat. Making the game longer, thereby making the connection between the first period and the continuation game more obvious, also has little impact on cooperation. For both the two- and three-period versions of our game we find substantial amounts of contingent play. This implies that use of a finitely repeated game, which has a number of methodological advantages, maintains the main properties of the underlying supergame.

JEL Classification: C91, C92, C73, L13

1. Introduction

The positive effect of allowing communication is possibly the strongest regularity in the literature studying cooperation in social dilemmas. Notable articles showing a positive effect from allowing communication include Dawes, MacTavish, and Shaklee (1977), Isaac and Walker (1988), Ostrom, Walker, and Gardner (1992), Charness and Dufwenberg (2006), Cason and Mui (2009), and Cooper and Kühn (2014; henceforth CK).[1] Communication has been shown to increase cooperation for both finitely and infinitely repeated games.

The primary purpose of this article is to explore two methodological questions that affect many studies of communication in games, particularly supergames. (i) The most common form of communication used by studies of communication in games (especially social dilemmas) is free-form messages or chat, but studies where subjects are limited to a narrow prespecified set of messages are far from rare.[2] Use of a limited message space has clear methodological advantages, allowing for clean tests of theoretical predictions and simplifying data analysis. Unfortunately, two existing articles (CK; Charness and Dufwenberg 2010) find that use of a limited message space almost eliminates the pro-cooperative effect of communication. In this article we explore

* Florida State University and University of East Anglia, 113 Collegiate Loop, Room 275, Tallahassee, FL 32306, USA; E-mail djcooper@fsu.edu; corresponding author (Contact information from January 2016 forward: University of East Anglia, DICE, and CEPR, Norwich Research Park, Norwich NR4 7TJ, UK)

† University of Michigan and CEPR, 611 Tappan Street, Ann Arbor, MI 48109, USA; E-mail kukuhn@umich.edu.

Received May 2014; accepted June 2015.

[1] See Sally (1995) for summary of findings from the psychology literature.

[2] Notable examples include Cooper et al. (1989 and 1992); Blume et al. (1998); and Blume and Ortmann (2007).

whether the seemingly negative effect of limiting the message space instead stems from a secondary feature shared by most limited message space designs – subjects are typically only allowed a single, simultaneous round of messaging. This makes it difficult for players to agree on a course of action, especially when the structure of the game is relatively complex. Will a limited message treatment still reduce cooperation relative to chat if it is possible to send multiple rounds of messages? (ii) The dominant methodology for studying supergames is to use an indefinitely repeated game (Roth and Murnighan 1978), but there are large benefits to using a finitely repeated game when communication is being studied (e.g., CK; Andersson and Wengström 2012). Because a finite game has a relatively small strategy space, the potential complexity of messages about what strategies should be used is limited, which makes analysis of chat data far easier. Use of a finite game also makes it easier to study learning and the effects of renegotiation. None of these advantages matter, however, if use of a finite game fundamentally changes subjects' behavior. If the relationship between stage games in a finitely repeated game is not transparent to subjects, dynamic incentives (i.e., the threat of punishment) could largely vanish (Frechette and Yuksel 2014). Will cooperation increase in a longer finite game that makes the connection between stage games more obvious, increasing dynamic incentives and cooperation?

To study these questions, we use the framework introduced in CK. CK studied a two-period game where cooperation can occur in equilibrium. The cooperative equilibrium is relatively complex, requiring contingent play where cooperation in the first period is supported by a threat to switch to a Pareto dominated equilibrium in the second period. CK explored why (rather than whether) communication increased cooperation. The relevant theory suggested that communication would increase cooperation by helping players coordinate on the cooperative equilibrium. However, there are many studies where communication improves cooperation even though cooperation is *not* consistent with equilibrium. This suggests that communication could operate through alternative channels such as subjects' preferences (i.e., increased group identity, lie aversion, guilt aversion) or improved understanding of the mutual benefits of cooperation. We were, therefore, particularly interested in the effect of threats to punish cheating since the relevant game theory suggested that this type of message should be important for supporting cooperation if the coordination channel is playing an important role. Consistent with the theory, threats of punishment were the most effective type of message for increasing cooperation.

After replicating a number of the results from CK, we present two new experiments. The first explores why the use of a limited message space seems to eliminate the positive effect of communication on cooperation. CK report lower levels of cooperation with a limited message space than with chat. This could be due to the use of a single, simultaneous round of messaging rather than the limited message space per se. Messages suggesting use of a contingent strategy (i.e., cooperate following cooperation and punish following cheating) are relatively complex and likely not obvious to many subjects. It is therefore rare that *both* subjects send such a message simultaneously. Suppose only one subject sends a message calling for cooperation supported by an incentive compatible punishment. For play of the proposed equilibrium to occur, the sender must believe the receiver understood the message and plans on following the suggestion. Likewise, the receiver must believe that the sender believes that the message was understood and will be followed. In other words, the subjects must reach an agreement. With a single, simultaneous round of messaging, there is no mechanism for reaching an agreement and communication is likely to have little effect even if the "right" message is sent.[3] Allowing for sequential sending of messages, especially with repetition, should make it easier for subjects to reach an agreement in a limited message treatment. If this yields similar choices to a chat treatment, experimenters could have the advantages that come with limited message space treatments while not fundamentally changing the effect of communication.

To examine this conjecture, we compare play with a single simultaneous round of messages to play with one or two rounds of sequential messages. Rates of agreement on either first period actions or full strategies (i.e., first period actions and second period actions *as a function of the first period outcome*) are much higher with sequential messages, but the overall levels of cooperation are unaffected. While subjects are reaching agreements more frequently, the proportion of messages that specify contingent play (i.e., punishment of cheating) in the second period is *not* increasing with sequential messages. The difference between the limited message space and chat treatments cannot be attributed to a failure to reach agreements, but rather to a failure to reach agreements *that make cooperation incentive compatible*. This reinforces our existing impression that communication via a limited message space is fundamentally different from chat. By extension, limited message space treatments may not be a good substitute for the use of chat to study communication.

The second new experiment explores the effect of using a three-stage game rather than a two-stage game. Unlike the two-stage game, the structure of the games in Periods 1 and 2 is identical (other than scaling) for the three-stage game. As such, the connection between the periods should be transparent. There is some evidence in favor of this. Dynamic incentives are always present, but, in the absence of communication, are stronger for the three-period game. This difference largely vanishes once communication is added, even with a limited message space that only allows subjects to specify a Period 1 choice but not any choices for Period 2 contingent on the Period 1 outcomes. More to the point, cooperation rates are largely unaffected by moving from two to three periods. Given the advantages of the two-period game for the study of communication, there is no overwhelming reason to move to longer finitely repeated games or indefinitely repeated games.

Both of our methodological findings have broad implications. Our first methodological result reinforces the existing evidence that communication via a limited message space is not equivalent to chat. The problem runs deeper than the standard use of a single round of simultaneous messaging. Allowing for more rounds of communication leads to more agreements about what course of action players should pursue, but ultimately has little effect on outcomes. The differing types of agreements with a limited message space versus chat suggests that use of a limited message space changes how subjects think about the game at a fairly deep level. Limited message space treatments have major advantages over chat, but the cost of changing subjects' thought processes may be greater than the benefit. Our second methodological result shows that making the connection between stage games more transparent by moving to a longer game has little impact on cooperation rates. For experiments looking at communication and cooperation in supergames, we feel that the advantages of the shorter games outweigh the disadvantages. We do *not* claim that the two-stage game (or finitely repeated games in general) is optimal for all studies of supergames. Indefinitely repeated games are closer to what most theorists have in mind for a supergame and, in

[3] The results of Cooper et al. (1989) are driven by a related problem. In a one-shot Battle of the Sexes game, two-way communication leads to significantly less coordination than one way communication. Because communication in one shot, subjects who initially disagree about what message should be sent have no means of reaching an agreement. Multiple rounds of play improve coordination.

the absence of communication, are likely to yield stronger dynamic incentives. For experimenters looking at equilibrium selection in supergames or the strategies used by subjects in supergames, there are good reasons for using an indefinitely repeated game and the benefits of the finitely repeated game are reduced. Different experimental questions call for different designs.

2. Methodological Issues

The most common method for studying communication in social dilemmas is free-form messages or chat, but experimental economic studies using limited message spaces are far from rare.[4] There are clear advantages to each method. Chat limits the possibility of demand induced effects, allows for unlimited back and forth between subjects, and does not restrict the message space. In contrast, with a limited message space subjects are shown a list of possible messages, suggesting these ought to be used. The effects of allowing communication may be measured incorrectly if the messages that matter most to subjects are not included in the message space or the lack of an opportunity to send further messages inhibits the ability of subjects to reach an agreement. On a more subtle level, the possibility of sending richer messages allows for more nuanced communication and may even stimulate thought about the problem at hand. CK note that performance is relatively poor with limited message spaces in part because subjects fail to specify incentive compatible punishments as part of agreements, as compared with common use of messages about incentive compatible punishment in the parallel chat treatment. The same types of messages about punishment are available in both treatments, but get used in a more sophisticated manner in the chat treatment. This suggests that subjects use a more sophisticated cognitive process in the chat treatment.

That said, limiting the message space allows for clean experimental identification of how specific messages affect behavior allowing for more precise tests of theories of communication. Chat content can be coded to perform similar exercises, but this is a complex exercise given the ambiguity of many messages and the presence of multiple relevant points in most conversations. If the primary advantage of laboratory experiments is the ability to exercise control over the decision-making environment, limited message spaces have an obvious advantage for studying the effect of communication.

This advantage vanishes if the choice to use a limited message space fundamentally changes the effect of communication. Existing evidence suggests this may be the case. Charness and Dufwenberg (2010) find that the effect of making promises in a trust game is reduced when subjects are limited to bare-bones promises and CK report greatly reduced rates of cooperation when subjects have only a limited message space.

Turning to the second methodological issue we faced in CK, the dominant methodology for studying supergames is to use an indefinitely repeated game, but there are substantial advantages to using a finite game. The set of available strategies in our two-stage game is far smaller than the set of strategies in the equivalent indefinitely repeated game. For experiments on communication, complexity is a fundamental problem. The set of messages needed in limited message space treatments, especially when contingent messages are allowed, quickly becomes too large to easily implement. Coding messages from chat sessions becomes far more complex as the set of possible

---

[4] Some notable examples include Cooper et al. (1989 and 1992); Blume et al. (1998); and Blume and Ortmann (2007).

	Low	Medium	High
Low	15	54	54
Medium	-24	45	114
High	-24	-24	60

Figure 1. Stage Game.

messages expands. With either a limited message space or chat, reaching an agreement is more difficult with a large set of possibilities to consider. Using a finite game limits the complexity of potential messages about what strategies should be used, reducing all of these problems.

The preceding was our primary reason for using a two-stage game in CK, but there are other advantages as well. Subjects' behavior changes as they gain experience playing a supergame against different opponents. This is true whether the supergame is implemented using an indefinitely repeated game (Dal Bó and Fréchette 2011) or a finite reduced form version of the game (CK). It follows that obtaining reliable results on cooperation in supergames, either with or without communication, relies on having achieved convergence. Repeated games with communication, especially chat, are slow, but the two-stage games are fast. Using the feedback meant that we could complete enough rounds to be reasonably sure of convergence even in treatments with multiple rounds of chat. Using a finite game also limits the complexity of the problem facing subjects as well as reducing the noise of feedback, both of which should speed learning by subjects.

One of the goals of CK was to examine the effects of renegotiation. Using a two-stage game simplified this task as the theoretical predictions of renegotiation proof equilibrium were simple and clean. With an indefinitely repeated game the theory becomes more complex with predictions varying depending on which version of renegotiation proof equilibrium is used. The two-stage game allows for a straightforward test of the relevant theory.

None of these advantages matter, however, if use of a finite game fundamentally changes subjects' behavior. This might occur if the two-stage game makes the connection between the two periods less transparent to subjects, reducing the role of dynamic incentives (see Frechette and Yuksel 2014).

3. Common Features of All Experiments

The Games

The basic stage game used in our experiments is shown in Figure 1. This can be interpreted as a standard Bertrand game in which firms have a choice between three possible prices and have a sunk cost of 24. The unique Nash equilibrium of the stage game played in isolation is (L,L). If this is played as an infinitely repeated game with a discount rate of $\delta = 2/3$, it is easily confirmed that mutual choice of High can be supported as a subgame perfect equilibrium.

Our experiments use two-period and three-period versions of this Bertrand game, referred to as the 2PBG and the 3PBG. For the 2PBG, the Period 2 stage game is derived from

Period 1

	Low	Medium	High
Low	15	54	54
Medium	-24	45	114
High	-24	-24	60

Period 2

	Low	Medium	High
Low	30	56	56
Medium	4	90	96
High	4	4	120

Figure 2. The Two-Period Bertrand Game (2PBG).

the continuation payoffs of the infinitely repeated version of the Period 1 stage game with a discount rate of $\delta = 2/3$. Specifically, we limit the players' available strategies to three versions of grim trigger and use pairings of these three strategies to generate a 3×3 payoff table (see Appendix C of CK for details of how this is done).[5] The resulting payoff table, shown on the right side of Figure 2, is a coordination game in which there are three pure strategy equilibria, (L,L), (M,M), and (H,H). These equilibria are Pareto ranked with (H,H) being the Pareto dominant equilibrium.

Considering the full two-period game, the necessary incentive conditions are satisfied so that (L,L), (M,M), or (H,H) can all be sustained in the first period of a subgame perfect equilibrium if players play (L,L) in Period 2 after any deviation in Period 1 and (H,H) otherwise. Asymmetric Period 1 outcomes (H,L) and (M,L) (and their permutations) can also be sustained in this way but not (H,M).[6] All collusive equilibria of the 2PBG, defined as equilibria that yield an outcome other than (L,L) for Period 1, require that players use the first period outcome as a coordination device for play in the second period. This captures the essential structure of all theories of collusion based on infinitely repeated games (Abreu 1998; Abreu, Pearce, and Stacchetti 1990) or finitely repeated games (Benoit and Krishna 1985).

Several other features of the 2PBG are worth noting. The equilibrium (M,M) is risk dominant (Harsanyi and Selten 1988) in the second period game. By not making (H,H) both Pareto and risk dominant, we make it less likely the second period play will gravitate so strongly to this outcome that contingent play does not occur. Not making (L,L) risk dominant makes it easier to identify play of Low in Period 2 as punishment rather than an attempt to play it safe. More generally, the 2PBG provides strong incentives in the sense that collusion at H is highly beneficial, strong punishments are needed to maintain collusion in equilibrium, and the loss from cheating and being punished in equilibrium is large. The payoff from play of (H,H) is 33% larger in each period than the payoff from play of (M,M), collusion at H in Period 1 can only be supported as an equilibrium outcome via reversion to (L,L) in Period 2, which reduces payoffs by 75% compared to play of (H,H), and the payoff from colluding at (H,H) in both periods is 25% greater than the payoff from defection to M in Period 1 followed by reversion to (L,L) in Period 2 (180

[5] The three versions of grim trigger strategy all choose an action for the first period (Low, Medium, or High) and repeat this action as long as the other player's action matches. If the other player ever chooses a different action, Low is played from that point forward. The "Low" version of grim trigger plays Low in all periods unconditionally.

[6] Note that (H,H) can only be sustained in period 1 if (L,L) is played in Period 2 after a deviation and (H,H) otherwise. All other Period 1 subgame perfect equilibria can also be sustained by playing (L,L) after a deviation and (M,M) otherwise in Period 2. The only Period 1 outcome that can be sustained by playing (M,M) after a deviation and (H,H) otherwise is (M,M) in Period 1.

vs. 144). Subjects are given strong incentives to reach and abide by collusive agreements, but the punishment called for by the equilibrium is sufficiently harsh for both players that it is unlikely to be undertaken lightly.

To generate Periods 2 and 3 of the 3PBG, the payoffs for Periods 1 and 2 of the 2PBG are discounted by 2/3. Period 1 of the 3PBG is given by the stage game payoffs shown in Figure 1.[7] The 3PBG has similar theoretical properties to the 2PBG. Specifically, mutual play of High in Period 1 can be supported as a subgame perfect equilibrium by the threat of mutual play of Low in Periods 2 and 3, but not by the threat of mutual play of Medium in Periods 2 and 3. Unlike the 2PBG, this is not the only way to support play of (H,H). For example, an even harsher punishment can be constructed by using the equilibrium in which the punisher plays Low and the player being punished plays Medium (or High) in Period 2.

Procedures

As much as possible, the procedures used for these new experiments were identical to those used for the original experiments reported in CK.[8] Subjects play 20 rounds of either the 2PBG or the 3PBG in all treatments. The design is between not within sessions, so subjects never switched between the two games. A "round" refers to an entire play of the 2PBG or 3PBG while a "period" refers to one of the two (or three) games played within a single round of the 2PBG (or 3PBG). Subjects are randomly matched with a new opponent in each round. Sessions are sufficiently large (minimum of 20 subjects) that it is unlikely that there are repeated game effects between rounds.

For the first 10 rounds in all treatments, subjects play the 2PBG or 3PBG without any communication. This is sufficient for subjects to understand the experimental interface, payoff tables, and main strategic issues in the 2PBG or 3PBG before attempting to master use of communication. Having 10 rounds of play before introducing communication also allows play to converge to the one shot Nash equilibrium in the first period, making the task facing subjects in Rounds 11–20 more challenging. Treatments vary by which game is being played, 2PBG or 3PBG, and the type of communication available in Rounds 11–20.

The experiments were conducted in the xs/fs laboratory at Florida State University (FSU). Subjects were recruited via ORSEE (Online Recruitment System for Economic Experiments; Greiner 2004). Sessions were run using z-Tree (Fischbacher 2007), and took between 1½ and 2 hours. Average earnings were slightly more than 25 dollars, including a 10-dollar show-up fee. Subjects were paid their total earnings from all 20 rounds. Payoffs were denominated in experimental currency units (ECUs) and were converted to dollars at a rate of 130 (390) ECUs equal $1 for the 2PBG (3PBG).

The instructions (see Supporting Information Appendix A) were read to the subjects, and were also shown on the subjects' computer screens. Several times the payoff tables were projected on an overhead screen for examples, making the payoffs common knowledge. The matching for this experiment is relatively complex (fixed matching within a round, random re-matching

---

[7] Discounting the payoff tables from the 2PBG by 2/3 results in fractions in the payoff table. To make the payoff tables easier for subjects to read, we avoided using fractions by multiplying all the "token" payoffs by three and then increasing the rate of token per dollar by a factor of 3. This leaves the dollar payoffs unchanged, but eliminates any fractions in the payoff table.

[8] Differences largely result from the experiments being conducted at FSU rather than CWRU. Notably, the mandated show-up fee at FSU ($10) was larger than the one used at CWRU ($6).

between rounds), so this point was emphasized. Following the instructions, subjects took a quiz testing their ability to read the payoff tables and their understanding of the instructions.

The experimental materials are framed using abstract language. For example, the materials do not refer to prices. Subjects choose between "A," "B," and "C" in Period 1, between "D," "E," and "F" in Period 2, and between "X," "Y," and "Z" in Period 3 (when relevant) with the three labels in each period corresponding to low, medium, and high prices. The terms "Low", "Medium", and "High" (or L, M, and H) are used throughout this article to ease exposition, but these are not the labels seen by subjects.

Subjects knew they would be playing a total of 20 rounds of the 2PBG. They also knew that the first ten rounds would be played without communication and followed by a pause for additional instructions. The possibility of communication was introduced at this intermediate point. To maintain parallelism, there was a pause prior to the 11th round in the treatment without communication to announce that none of the rules would change.

In treatments with limited message spaces, the instructions prior to round 11 described in detail what messages could be sent, but provided no guidance about why any particular message should be sent. We stressed that the messages are cheap talk with no direct effect on payoffs.

Subjects in the chat treatment received extensive instructions, largely focused on the mechanics of using the chat program. The instructions gave the subjects no guidance on what types of messages should be sent other than (i) requesting that they not identify themselves and (ii) asking them to avoid offensive language. These instructions also stressed that the messages are cheap talk.

Subjects had printed copies of the payoff tables for *all* periods available whenever they made a decision. When choosing a price, the interface showed the outcomes for previous periods (prices and payoffs) as well as a summary of outcomes for all previous rounds. The interface automatically showed the summary for the three most recent rounds with a scroll bar that could be used to see earlier rounds. Feedback at the end of each round also included the sum of payoffs across all periods for both players as well as the period-by-period outcomes. Subjects did not receive identifying information about their opponent, such as a subject ID number, to limit the possibility of repeated game effects across rounds.

4. Replication

The treatments in CK differ in whether communication is allowed prior to Period 1, what type of communication is allowed, and whether renegotiation is possible. We replicated four of the five treatments (see fn. 13 for an explanation of why the fifth treatment was not replicated). A summary of these four treatments follows. Subjects play 20 rounds of the 2PBG in all treatments. Recall that a "round" refers to an entire play of the two-period game while a "period" refers to one of the two stage games played within a single round of the 2PBG. For the first 10 rounds in all treatments, subjects play the 2PBG without any communication, so treatments only vary the type of communication that is possible in Rounds 11–20.

1. *No Communication (N Treatment)*: The rules for Rounds 11–20 are identical to those in Rounds 1–10, with no communication between players.
2. *Period 1 Limited Communication (P1 Treatment)*: The P1 treatment gives subjects the opportunity to send a message prior to the beginning of Period 1. The message

Table 1. Summary of Treatments for Replication

	N		P1		P1C		Chat	
	FSU	CWRU	FSU	CWRU	FSU	CWRU	FSU	CWRU
# Sessions	2	3	2	3	4	3	2	3
# Subjects	46	64	48	68	94	74	48	64
Period 1 limited messages			✓		✓			
Contingent messages					✓			
Period 1 chat							✓	

space is limited to suggesting actions for Period 1. Specifically, subjects are given the prompt, “I think we should choose the following in Period 1.” They are asked to choose between “Low,” “Medium,” “High,” or “No Response” for both “My Choice” and “Your Choice.” Messages are chosen simultaneously and each player is shown both parts of both players' messages at the same time as choices are made for Period 1.[9] The feedback at the end of Period 1 reiterates the messages as well as reporting the outcome for Period 1. Subjects cannot send any messages about their intent for Period 2.

3. *Period 1 Limited Communication with Contingencies (P1C Treatment)*: In addition to specifying what actions should be chosen for Period 1, subjects in the P1C treatment also indicate what actions should be chosen for Period 2 subject to the outcome for Period 1. The set of possible messages about Period 2 is limited: subjects are prompted “[i]f we choose the preceding [i.e. the actions the subject has specified should be chosen for Period 1] in Period 1, I think we should choose the following in Period 2” and “[i]f we DO NOT choose the preceding [i.e. the actions the subject has specified should be chosen for Period 1] in Period 1, I think we should choose the following in Period 2.” Limiting the message space simplifies the problem facing subjects while still making it possible to send a credible message that supports collusion in Period 1. Subjects are shown both players' messages, in full, at the same time as choices are made for Period 1. The messages are displayed again as part of the feedback for Period 1.
4. *Pre-Play Chat (Chat Treatment)*: Starting in Round 11, players can communicate prior to Period 1 using the chat option in z-tree (Fischbacher 2007). This is very similar to using an instant messaging program. Continuous back-and-forth communication is possible until one of the players makes a decision for the period.

Table 1 summarizes the data gathered for the replications study as well as the original data gathered for CK at Case Western Reserve University (CWRU). CK gathered three sessions for all of their treatments. For Experiment 1 (see section 5), we gathered three or four sessions per treatment. The P1C data for FSU comes from this experiment. In Experiment 2 (see section 6), we gathered two sessions. The N, P1, and Chat data for FSU is from Experiment 2. See Supporting Information Appendix B for a detailed summary of the FSU data.

Figure 3 shows data from CWRU in the left panel and parallel data from the FSU replication in the right panel. The fraction of players choosing High is shown on a round-by-round basis for

[9] For 95% of messages across all limited message treatments, subjects choose the same messages for both players (i.e., always called for symmetric play).

each of the four treatments. Choice of High is a natural measure of cooperative play, although our conclusions do not depend on the details of how cooperative play is defined.

CK reported a clear pattern of treatment effects. By Round 10, cooperation has collapsed in all treatments with virtually no remaining play of High. In the N treatment, cooperation remains almost nonexistent throughout Rounds 11–20. In the two treatments with limited message spaces, there is an initial jump in cooperation (choice of High) when messaging is introduced. Cooperation is consistently higher with contingent messages about Period 2 choices (P1C) than without (P1), but cooperation fades over time in both treatments. The initial jump in cooperation is larger with chat, this treatment always produces the highest levels of cooperation, and cooperation does not disappear with experience.

The same pattern of treatment effects is observed in the FSU data. Cooperation has largely vanished by Round 10, with 85% of subjects choosing Low for Period 1. Virtually no cooperation is observed in Rounds 11–20 without communication. Cooperation jumps in both limited message space treatments for Round 11, more so for P1C than P1, but collapses over time. With chat, the initial increase in cooperation is larger and cooperation remains steady over time.[10]

Result 1: The treatment effects reported by CK are replicated in the FSU data.

While the overall pattern of results is similar in the CWRU and FSU data, there are differences worth noting. For all three treatments with communication, the levels of cooperation are higher in the FSU data than in the CWRU data. The difference between cooperation rates in the two locations is statistically significant at the 1% level.[11] This difference is also present in Rounds 1–10 prior to the introduction of communication. Choice of High is not a useful metric in the early rounds since it is universally rare, but using an expanded metric of cooperation, choice of Medium or High, was more frequent for Rounds 1–10 in FSU (38%) than in CWRU (28%). Once again the difference between CWRU and FSU is significant at the 1% level.[12]

We have no explanation for why the FSU population is consistently more cooperative than the CWRU population. The two populations differ on multiple dimensions that could explain the difference. CWRU is a highly selective, expensive, private university in the Midwest while FSU is a less-selective, inexpensive, public university in the South. There is no obvious feature in the data that provides an explanation either. For instance, looking at P1 and P1C, pairs that agree on mutual play of High in Period 1 tend to cooperate more. You might, therefore, suspect that

[10] To confirm that differences between treatments were significant in the FSU data, we ran an ordered probit regression using FSU data from Rounds 11–20. The dependent variable is a subject's Period 1 choice (0 = Low, 1 = Medium, 2 = High) and independent variables were treatment dummies and round dummies. Standard errors are corrected for clustering at the subject level. The differences between the N and P1 treatments ($z = 8.97$), P1 and P1C treatments ($z = 3.79$), and P1C and PChat treatments ($z = 9.10$) are all significant at the 1% level. These results are little affected by including a control for the subject's behavior in Rounds 1–10.

[11] To reach this conclusion, we ran an ordered probit regression using data from Rounds 11–20. The dependent variable is a subject's Period 1 choice (0 = Low, 1 = Medium, 2 = High) and independent variables were treatment dummies, round dummies, and a dummy for FSU. Standard errors are corrected for clustering at the subject level.

The FSU dummy is significant at the 1% level ($z = 5.30$). The results are little affected by including a control for the subject's behavior in Rounds 1–10. The difference between locations is significant for each of the three treatments with communication when considered in isolation.

[12] Running an ordered probit with rounds dummies and an FSU dummy, the FSU dummy is significant at the 1% level ($z = 3.44$).

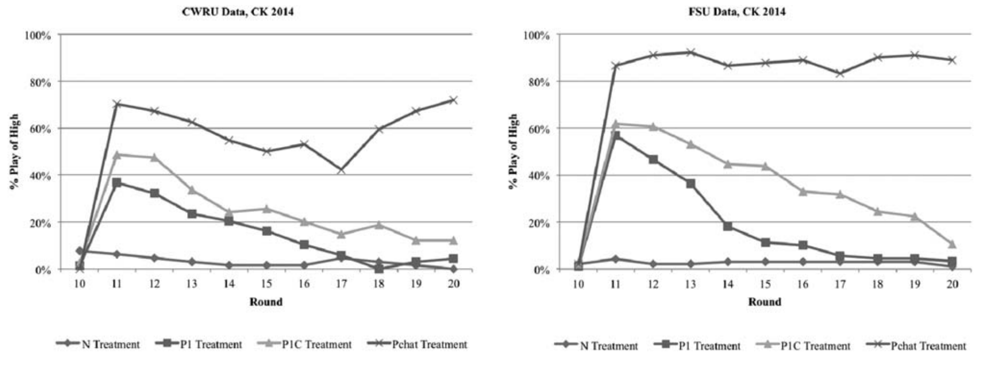

Figure 3. Replication of Results from CK, 2014.

subjects in FSU are more likely to reach an agreement or are more sensitive to agreements. You would be wrong, as subjects at CWRU were more likely to reach an agreement than those at FSU (50% vs. 36%) and were more responsive to agreements.

Higher cooperation rates in the FSU data are most pronounced for the Chat treatment. In the CWRU data, there is an initial period of declining cooperation in the Chat treatment followed by a recovery. This pattern is absent in the FSU data as choice of High is consistently extremely frequent in the Chat treatment.[13]

5. Experiment 1 – Simultaneous versus Sequential Message

The P1C treatment allows subjects to send credible messages supporting mutual play of High in Period 1, as reversion from mutual play of High to mutual play of Low in Period 2 provides sufficient incentives to support Period 1 cooperation. As can be seen in section 3, this possibility was not sufficient to support long-term cooperation in either the CK data or the new data gathered at FSU. This could represent a fundamental difference between chat and limited message space treatments, or could be a purely mechanical function of how the limited message space treatment was implemented. An important function of sending messages is coordination on a particular equilibrium, but a single round of simultaneous messages may not be sufficient to build the necessary common knowledge of an equilibrium. Chat is inherently sequential in nature and makes it easy to build common knowledge through agreements. If I send you a message proposing play of an equilibrium, you can send back a message agreeing. Going further, I can even reiterate my proposal and you can reiterate agreement. The P1C treatment might have been more successful if subjects had a better opportunity to reach agreements. Experiment 1 explores this conjecture by introducing two new treatments.

Subjects play 20 rounds of the 2PBG in all treatments of Experiment 1. For the first 10 rounds in all treatments, subjects play the 2PBG without any communication, so treatments only

[13] CK included a fifth treatment (RChat) that allowed for renegotiation, including chat both prior to Period 1 and between Periods 1 and 2. CK find higher cooperation in the RChat treatment than in the Chat treatment. We did not attempt to replicate this result. Given that the Chat treatment already produced near perfect cooperation at FSU, higher than was observed even for the RChat treatment at CWRU, there was no room for cooperation to be higher with renegotiation. We, therefore, saw no reason to spend money and time adding an RChat treatment at FSU.

Table 2. Summary of Treatments for Experiment 1

	P1C	Seq1	Seq2
# Sessions	4	3	3
# Subjects	94	71	63
Period 1 limited messages	✓	✓	✓
Contingent messages	✓	✓	✓
Sequential messages		✓	✓
Two rounds of messages			✓

vary the type of communication that is possible in Rounds 11–20. The two new treatments are as follows.

One Round of Sequential Messages (Seq1): The message space is identical to P1C, but the two players send messages sequentially, with the second player observing the first player's message before choosing a message. The order in which players send messages is selected randomly. Because the second player sees the first player's message before choosing their own message, they can implicitly agree with the first player by sending an identical suggestion.

Two Rounds of Sequential Messages (Seq2): The message space is identical to P1C, but the two players send *two* rounds of *sequential* messages. The order in which players send messages is selected randomly and remains the same for both rounds of messaging. Subjects observe previous messaging before choosing a message. More than one round of messaging should make it easier to reach an agreement. For example, rather than having to agree or disagree with the first player's proposal, the second player now has the option to make a counter-proposal that the first player can agree to.

Table 2 summarizes the data gathered for Experiment 1.[14] We hypothesized that cooperation would be more frequent with sequential messages, and more frequent with two rather than one round of messaging. Underlying these predictions was an assumption that more and sequential messaging would improve the ability of players to reach agreements, both on Period 1 actions and on full strategies.

Figure 4 summarizes choice of High in Period 1 for the three treatments in Experiment 1. There is little difference between the treatments. In all three, there is an initial increase in choice of High when messages are introduced for Round 11, and in all three the initial burst of cooperation dissipates over time. There are no statistically significant differences in Period 1 choices between the three treatments in Experiment 1.[15]

Our initial hypothesis was that sequential messaging would increase cooperation by increasing the likelihood that players reached an agreement. The failure to see a treatment effect on cooperation could reflect a failure to increase the frequency of agreements. Figure 5 makes it clear that this is not the case. The left panel shows the percentage of pairs that reached an agreement on Period 1 choices. The right panel uses a stronger definition of agreement, requiring agreement on

[14] Data from two subjects have been dropped. One of these subjects had participated in an earlier session and the other was visibly (and pungently) stoned.

[15] To reach this conclusion, we ran an ordered probit regression using data from Rounds 11–20 of Experiment 1. The dependent variable is a subject's Period 1 choice (0 = Low, 1 = Medium, 2 = High) and independent variables were treatment dummies and round dummies. Standard errors are corrected for clustering at the subject level. None of the differences between treatments are statistically significant at the 10% level. The largest difference is between P1C and Seq2, but even this difference does not approach significant ($z = 0.83$). The results are little affected by including a control for the subject's behavior in Rounds 1–10.

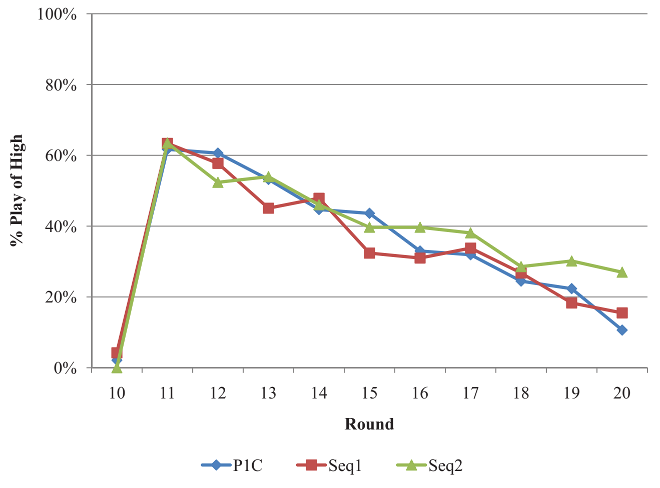

Figure 4. Simultaneous versus Sequential Messaging.

full strategies: Period 1 choices and Period 2 choices *contingent on the outcome of Period 1*. For either definition, it is clear that agreements were more common with sequential messages, especially when there were two rounds of sequential messages.

The problem is not reaching an agreement, but rather what messages are sent and what agreements are reached. Contingent messages are effective. When a pair of subjects agree to both choose High in Period 1, the probability that an individual follows through by choosing High in Period 1 jumps from 44% to 70% if the final message sent by one of the subjects is contingent (i.e., calls for differing actions in Period 2 depending on Period 1 outcomes) and reaches 93% if both subjects used contingent messages. Unfortunately, while the frequency of agreements rises with sequential messages (and even more with two rounds of sequential messages), the frequency of games where a contingent message is sent falls from 33% in P1C to 28% in Seq1 and a mere 17% in Seq2. Out of 1140 games played in Rounds 11–20, there were only 17 games where the pair of subjects agreed on a contingent strategy (calling for differing actions in Period 2 depending on Period 1 outcomes). Subjects may be reaching more agreements in Seq1 and Seq2, but they almost always agree on mutual play of High in Period 2 regardless of the Period 1 outcome (96% of agreements on full strategies). This does not yield the incentives needed to support cooperation in Period 1.

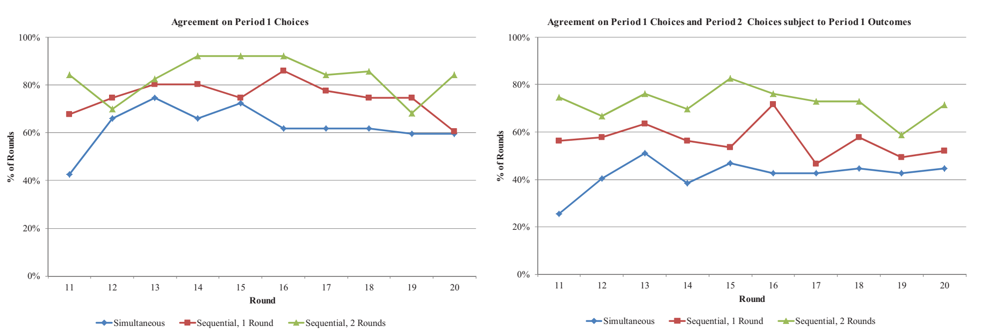

Figure 5. Agreements, Simultaneous versus Sequential.

It is worth noting that failure to send contingent messages does not imply that Period 2 choices are unaffected by Period 1 outcomes. In games where the two players agree to play High unconditionally in both periods, the most common type of agreement, the proportion of players choosing High in Period 2 falls from 82% when the other player chooses High in Period 1 to 56% if he chooses Medium and 45% if he chooses Low. Subjects often punish opponents who break agreements, but they fail to grasp the value of warning their opponents that cheating will be punished.

Result 2: Allowing sequential messages or increasing the rounds of sequential messages increases the likelihood that subjects reach an agreement but does not significantly affect the likelihood that High is chosen in Period 1.

6. Experiment 2 – Two- versus Three-Period Game

As noted previously, there are clear methodological advantages to using two-period games in an experiment that focuses on communication. However, there are potential costs as well. If subjects fail to perceive a link between the Bertrand game played in Period 1 and the coordination game played in Period 2, this could lower the incentive to choose High in Period 1 and, therefore, systematically reduce cooperation. Our concern is not so much that play should be identical in the two-period game versus longer games, but that use of a two-period game might either affect the treatment effects or, at a deeper level, change the nature of the strategic problem facing subjects.

Experiment 2 explores this by comparing behavior in the 2PBG and the 3PBG. While both games share the common features of being finite games that terminate in a coordination game and rely on punishment in the coordination game to support equilibrium choice of High in Period 1, the link between Period 1 and the continuation game should be clearer in the 3PBG since the games are identical (subject to discounting) in Periods 1 and 2. If this increases the incentives to cooperate in Period 1, we expect to see more cooperation in the 3PBG.

We compared the 2PBG and the 3PBG for three treatments: no communication (N), Period 1 limited communication (P1), and pre-play chat (Chat). These three treatments allow for identical messaging between the two games. We did not run the 3PBG for the P1C treatment because it was not obvious what would be the correct message space. If we wanted to follow the intent of P1C, we would have allowed subjects to specify actions for Period 2 and Period 3 contingent on the Period 2 outcome. This yields an enormous message space which would be difficult to implement as part of a limited message space treatment. The message space can be limited by only allowing subjects to specify Period 2 actions, but this implies that subjects can agree on a full equilibrium (including all contingencies) in the 2PBG but not the 3PBG. From a theoretical point of view, we would no longer expect the two games to have identical properties.

Two sessions were run for each treatment. Other than the change from two to three periods per game, the procedures are identical in 2PBG and 3PBG sessions.

Figure 6 shows the percentage of subjects choosing High in Period 1 as a function of the game (2PBG or 3PBG) and treatment (N, P1, or Chat). While there are small differences by treatment, overall the move from the 2PBG to the 3PBG does not have a dramatic effect on the rate of choosing High in Period 1. Cooperation dies out over Rounds 1–10 in both cases. In Period 1 of Round 10, 90% of subjects choose Low in the 2PBG versus 92% in the 3PBG. The treatment with

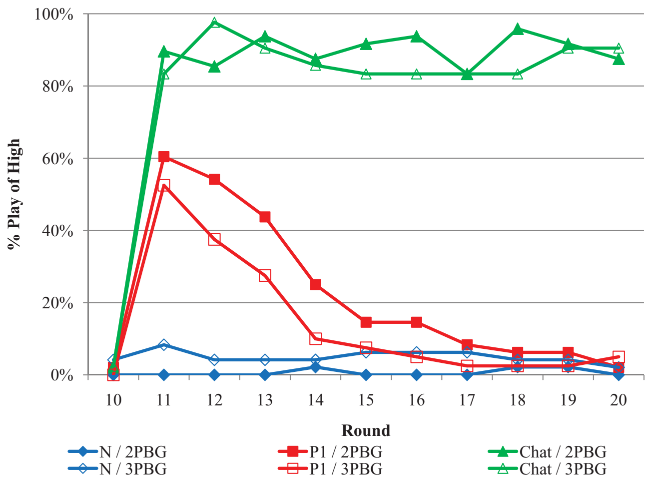

Figure 6. Two- versus Three-Period Games.

the clearest effect in Rounds 11–20 is P1, but moving to the 3PBG reduces cooperation rather than increasing it.

Pooling across treatments, the effect on Period 1 choices of moving from the 2PBG to the 3PBG is not statistically significant.[16] Breaking down the effect by treatments, the difference between the 2PBG and 3PBG is significant at the 5% level for P1 and at the 10% level for the N treatment.[17] These results should be taken with a grain of salt; Experiment 2 was designed to compare the two games across treatments, and so there is not a huge amount of data for any one treatment (the six cells range between 40 and 48 subjects). Given that there is not a significant difference between the 2PBG and 3PBG in Rounds 1–10,[18] when much more data is available and play is not compacted against a boundary, we doubt that there is truly a difference between the two games in the N treatment.

Result 3: Use of the 3PBG has only a minor impact on Period 1 choices relative to the 2PBG, and does not seem to affect the differences between treatments.

The data from Experiment 2 also allows us to address the issue of whether using of the 2PBG reduces contingent play (i.e., changes in Period 2 play as a function of Period 1 play). Frechette and Yuksel (2013) report data from a related treatment. This is a five-period game with four periods of a repeated prisoner's dilemma followed by a fifth period coordination game derived from continuation payoffs. They report weak dynamic incentives, as they find little relationship between

[16] To reach this conclusion, we ran an ordered probit regression using data from Rounds 11–20 of Experiment 2. The dependent variable is a subject's Period 1 choice (0 = Low, 1 = Medium, 2 = High) and independent variables were treatment dummies, round dummies, and a dummy for the 3PBG. Standard errors are corrected for clustering at the subject level. The dummy for the 3PBG is not statistically significant ($z = 1.14$). This result is unaffected by including a control for the subject's behavior in Rounds 1–10.

[17] The ordered probits described in fn. 16 are modified by including interactions between the treatment dummies and the dummy for the 3PBG. This is the one case where controlling for cooperation rates in Rounds 1–10 affects the results, reducing the significance of the 3PBG dummy for P1 from the 1% level to the 5% level ($z = 2.08$). The 3PBG dummy is significant at the 10% level in either case for the N treatment ($z = 1.88$ with control for Round 1–10 cooperation).

[18] Running the appropriate ordered probit, the dummy for the 3PBG is not significant ($z = 0.97$).

Table 3. Average of Decision 2 (Low = 0, Medium = 1, High = 2)

	Rounds 1–10		Rounds 11–20	
Other's Decision 1	2PBG	3PBG	2PBG	3PBG
Low	1.07	0.22	0.91	0.03
Medium	1.24	0.41	1.17	0.13
High	1.46	0.85	1.88	1.59

what happened in the Period 1 prisoner's dilemma and what happens in the Period 5 coordination game. We examine a slightly different question which we think is more germane given our focus on Period 1 play. For our use of a two-period game to be appropriate, there should be clear dynamic incentives—Period 2 play should depend on Period 1 outcomes. While dynamic incentives are lower in the 2PBG relative to the 3PBG, contingent play is clearly present in both treatments.

To see this point, Table 3 shows average decisions for Period 2 (Low = 0, Medium = 1, High- = 2) as a function of the other player's Period 1 decision. The data is broken down between Rounds 1–10, when no communication is available, and Rounds 11–20 when communication is possible (P1 and Chat treatments). Only data from the P1 and Chat treatments is being included for Rounds 1–10 so the same subjects are being compared in the two time blocks.[19] Data within each time block are broken down between the 2PBG and 3PBG.

Across the board, Period 2 decisions respond to the other player's Period 1 outcomes. There is no shortage of contingent play. Contingent play is stronger with communication, although this might be an order effect as contingent play generally grows over time. Comparing the 2PBG and 3PBG, Period 2 choices are consistently more sensitive to the other player's Period 1 choice in the 3PBG than in the 2PBG.

The preceding points were made using a specific measure and data set, but are robust to use of other measures (i.e. continuation payoffs, relationship between Period 1 and Period 3 in the 3PBG) or data sets (i.e., broadening the data set to include all treatments from Experiment 2 or all treatments from Experiments 1 and 2). There is no shortage of contingent behavior.

To make the preceding point more formally, we ran ordered probit regressions. All data is drawn from the P1 and Chat treatments, and separate regressions were run for Rounds 1 – 10 and Rounds 11–20. The dependent variable is a player's Period 2 choice (Low = 0, Medium = 1, High- = 2). The independent variables are round dummies, a dummy for the chat treatment (in Rounds 11–20), the other player's Period 1 choice (Low = 0, Medium = 1, High = 2), a dummy for the 3PBG, an interaction between the other player's Period 1 choice and the chat dummy (in Rounds 11–20), and an interaction between the other player's Period 1 choice and the 3PBG dummy. Table 4 reports results for the parameters of primary interest, the other player's Period 1 choice and the interaction between this and the 3PBG dummy. Full regression output is available from the authors upon request. Both regressions are based on 1780 observations from 178 subjects. Standard errors are clustered at the subject level.

All parameters reported in Table 4 are statistically significant at the 1% level. In other words, significant contingent play is always observed in both the 2PBG and 3PBG, but the response of

[19] The data from the N treatment in Rounds 11–20 are not useful for making comparisons between the 2PBG and 3PBG because there is virtually no play of Period 1 actions other than Low.

Table 4. Ordered Probit Regressions Comparing Contingent Play in the 2PBG and 3PBG

	Rounds 1–10		Rounds 11–20	
Independent Variable	Parameter	St. Error	Parameter	St. Error
Other's Period 1 choice	0.220	0.075	0.326	0.078
3PBG * Other's Period 1 choice	0.317	0.107	0.793	0.210

Period 2 play to the other player's Period 1 choice is significantly stronger in the 3PBG. Note that we are reporting parameter estimates, not marginal effects. Contingent play is significantly stronger, but not almost three times stronger in the 3PBG for Rounds 11–20. The other player's Period 1 choice is arguably endogenous for Rounds 11–20 due to communication. To address this issue, we also ran a two-step procedure where the other player's Period 1 choices for the three preceding rounds were used as an instrument for the current Period 1 choice. This does not affect our conclusion that there is significant contingent play for both the 2PBG and 3PBG, with contingent play being significantly stronger in the 3PBG.

RESULT 4: There is significant contingent play for both the 2PBG and 3PBG, with contingent play being significantly stronger in the 3PBG.

7. Conclusion

Using new data, we replicate the differing cooperation rates observed across treatments in CK (with the exception of the RChat treatment that was not run at FSU). Choice of High in Period 1 is lowest with no communication, and increases with the introduction of limited messages about Period 1 actions (P1), limited messages about actions in Period 2 contingent on the Period 1 outcome (P1C), and chat prior to Period 1 (Chat). Only with chat are high stable levels of cooperation observed throughout Periods 11–20. This is more cooperation than observed with chat in CK, where Period 1 choice of High collapses and then recovers to high levels.[20]

Experiment 1 confirms that the collapse of cooperation in P1C, the limited message space treatment with contingent messages, is not driven by an inability of subjects to reach agreements due to only having a single, simultaneous round of messaging. Even when we make it easier to reach an agreement, subjects tend to reach agreements that do not make Period 1 cooperation incentive compatible.

Experiment 2 looks at whether cooperation increases when we make the relationship between stage games more transparent by moving from the 2PBG to the 3PBG. We find little evidence to suggest this. We find contingent play in both the 2PBG and 3PBG, but contingent play is weaker in the 2PBG. This is related to but less extreme than the finding of Frechette and Yuksel—the relationship between play in Periods 1 and 2 is weaker when Period 2 is the coordination game rather than a discounted version of the Period 1 game, but strong dynamic incentives are still present.

Experiment 1 does not imply that communication via a limited message space could *never* be as effective as chat, but instead eliminates one obvious explanation for its weak performance in

[20] CK relate this recovery to the increased use of messages specifying punishment. Messages threatening punishment also increase over time in the current PChat data set, but obviously this cannot be related to a recovery in cooperation rates since Period 1 cooperation is high throughout.

previous experiments. There are many types of messages in the chat treatment that are not available in any of the limited message treatments, and it is possible that if sufficient messages were added to the message space, high degrees of cooperation could also be achieved with a limited space. This somewhat reduces the primary methodological advantage of using a limited message space, the relative simplicity of the messages that need to be analyzed, and does not eliminate the primary source of concern. If use of a limited message space can fundamentally change how subjects think about a game, the benefits of using limited message spaces may not be worth the costs.

The results of Experiment 2 need to be interpreted with caution. Play is not identical across the 2PBG and 3PBG, but there is ample contingent play in both cases and the differences are not likely to be driving our conclusions about treatment effects or the conclusions reached in CK. It does not follow that use of the 2PBG is harmless. For studies examining communication, like CK, there are major advantages to have a game that is fast and has a limited message space. The effects of using such a short game do not matter much for the treatment effects of interest. For experiments studying other issues about supergames, this would not be true. For instance, we would *not* recommend using the 2PBG (or a similar game) to study the effect of changing the discount rate on equilibrium selection. It also must be noted that the 2PBG was designed carefully so the coordination game in Period 2 would respond to Period 1 outcomes. Contingent play in the 2PBG cannot be taken for granted. It requires careful engineering of the payoff table in ways that may not be appropriate for some studies.

Acknowledgments

We would like to thank the National Science Foundation (SES-0720993) for funding these experiments. Sean Collins, E. Glenn Dutcher, John Jensenius, Cortney Rodet, and Krista Jabs Saral provided valuable research assistance on this project. Valuable feedback was provided by the editor, Cary Deck, and two anonymous referees. The authors are solely responsible for any errors in this manuscript.

References

Abreu, Dilip. 1998. On the theory of infinitely repeated games with discounting. *Econometrica* 56:383–96.

Abreu, Dilip, David Pearce, and Ennio Stacchetti. 1990. Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica* 58:1041–63.

Benoit, Jean-Pierre, and Vijay Krishna. 1985. Finitely repeated games. *Econometrica* 53:905–22.

Blume, Andreas, Douglas V. DeJong, Yong-Gwan Kim, and Geoffrey B. Sprinkle. 1998. Experimental evidence on the evolution of meaning of messages in sender-receiver games. *American Economic Review* 88:1323–40.

Blume, Andreas, and Andreas Ortmann. 2007. The effects of costless pre-play communication: Experimental evidence from games with Pareto-ranked equilibria. *Journal of Economic Theory* 132:274–90.

Cason, Tim, and Vai-Lam Mui. 2009. Coordinating resistance through communication and repeated interaction. Working paper, Purdue and Monash Universities.

Cason, Tim, and Vai-Lam Mui. 2014. Coordinating resistance through communication and repeated interaction. *The Economic Journal* 124:F226–F256.

Charness, Gary, and Martin Dufwenberg. 2006. Promises and partnership. *Econometrica* 74:1579–601.

Charness, Gary, and Martin Dufwenberg. 2010. Bare promises. *Economic Letters* 107:281–83.

Cooper, Russell, Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross. 1989. Communication in the battle of the sexes game: Some experimental results. *Rand Journal of Economics* 20:568–87.

Cooper, Russell, Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross. 1992. Communication in coordination games. *The Quarterly Journal of Economics* 107:739–71.

Cooper, David J., and Kai-Uwe Kühn. 2014. Communication, renegotiation, and the scope for collusion. *American Economic Journal: Microeconomics* 6:247–78.

Dal Bó, Pedro, and Guillaume R. Fréchette. 2011. The evolution of cooperation in infinitely repeated games. *The American Economic Review* 101:411–29.

Dawes, Robyn M., Jeanne McTavish, and Harriet Shaklee. 1977. Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of Personality and Social Psychology* 35:1–11.

Fischbacher, Urs. 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10:171–78.

Frechette, Guillaume, and Sevki Yuksal. 2013. Infinitely repeated games in the laboratory: Four perspectives on discounting and random termination. Working paper, New York University.

Greiner, Ben. 2004. An Online Recruitment System for Economic Experiments. In *Forschung und wissenschaftliches Rechnen* edited by Kurt Kremer and Volker Macho. GWDG Bericht 63, Göttingen: Ges. für Wiss. Datenverarbeitung, pp. 79–93.

Harsanyi, John, and Reinhard Selten. 1988. *A general theory of equilibrium selection in games*. Cambridge: MIT Press.

Isaac, Mark R., and James M. Walker. 1988. Communication and free-riding behavior: The voluntary contribution mechanism. *Economic Inquiry* 26:585–608.

Ostrom, Elinor, James M. Walker, and Roy Gardner. 1992. Covenants with and without a sword: Self-governance is possible. *American Political Science Review* 86:404–17.

Roth, Alvin E., and J. Keith Murnighan. 1978. Equilibrium behavior and repeated play in prisoners' dilemma games. *Journal of Mathematical Psychology* 17:189–98.

Sally, David. 1995. Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society* 7:58–92.

Van Huyck, John B., Raymond C. Battalio, and Richard O. Beil. 1990. Tacit coordination games, strategic uncertainty, and coordination failure. *The American Economic Review* 80:234–48.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publishers web-site:

Supporting Information Appendix Table A4.