Communication, Renegotiation, and the Scope for Collusion

David Cooper
Florida State University

Kai-Uwe Kühn
University of Michigan and CEPR

April 2009
First Draft – Preliminary and Incomplete

Abstract: In this paper we experimentally analyze the type of communication that is needed to achieve persistent collusion in a simple collusion game. As expected from earlier literature on communication and collusion, any type of communication leads to a short run treatment effect. However, consistent with theories of credible cheap talk, subjects learn very quickly that communication is not credible when the complete contingent strategies cannot be conveyed. In a treatment with unrestricted pre-game communication via a chat window we can show that communication of a contingent strategy threatening punishment for non-compliance is the most successful strategy to achieve collusion. Furthermore, subjects learn through repeated play to use such strategies, which helps the establishment of a positive treatment effect in the long run. Surprisingly, collusion is even more successful when we allow for renegotiation. This is despite the fact that individuals almost always renegotiate to the Pareto optimal continuation equilibrium and therefore face no punishments for deviations from first period collusion. However, we show that non-monetary punishments in form of verbal admonishment in the renegotiation phase is sufficient to generate incentives to stick to collusive promises in period 1. Renegotiation therefore does not just eliminate punishment in the market but also creates a new punishment channel via non-monetary payoffs. The paper tests for the presence of a variety of other behavioral explanations for collusive outcomes.

1. Introduction

Collusion theory has become a standard workhorse in antitrust policy advice. The theory of repeated games (see Abreu 1988, Abreu, Pearce, Stacchetti) has provided a canonical framework for the analysis of collusion problems. However, in policy applications it is often ignored that the basic theory relies on the exploitation of a coordination problem in games with many equilbiria. In particular, players have to make continuation play contingent on the outcome of previous periods in a coordinated way. It has often been assumed that such coordination is easily achievable and as a result tacit and explicit collusion have often been treated equivalently in the economics literature. However, we now know that even in much simpler coordination games experimental subjects often dramatically fail to achieve Pareto optimal outcomes in the absence of an explicit coordination device.

In this paper, we analyze the impact of communication as a coordination device in an experimental collusion game. Our study of communication allows us to test two aspects of theory at the same. First, it allows us to test whether theories of credible pre-game communication as discussed by Farrell and Rabin (19??) can predict the type of communication we observe in our experimental collusion games. If these theories have predictive power, they can also help us to test the underlying collusion theory experimentally. Without communication in experiments we can only observe actual play, but not whether and what contingent strategies are actually chosen by the players. According to the theories of credible pre-play communication discussed by Farrell and Rabin it is, however, necessary to communicate the complete contingent strategy between players in order to achieve collusive results.

We show in this paper that the theories of pre-play communication have low predictive power for the short run treatment effect of commuincation, but have very strong predictive power for the long run treatment effect, i.e. for experienced subjects. First, when communication is limited to only communication about collusive actions but contingencies cannot be conveyed, there is a significant treatment effect. However, the impact of such communication quickly disappears and communication looses all information content in the long run. This is fully consistent with the the theories of pre-game communication and also replicates earlier results in collusion experiments with communication that have not allowed for contingent statements. Our innovation to the literature comes from allowing for unlimited communication via a chat function. We show that the sepcification of contingent strategies is the most successful communication strategy for establishing collusive outcomes. Furthermore, we show that only few players initiate contingent strategies, but others learn to play contingent strategies in the long run achieving a high level of collusion. Again this confirms both the theories of pre-play communication and the basic structure of game theoretic collusion models.

Our data also suggest that just making contingent statements is not enough to achieve high levels of collusion, but that rich language and explanation of the strategy suggested is important for subjects to learn contingent play. When only very structured contingent statements are allowed, collusive levels greatly deterioriate over time. However, collusion does not completely disappear even when contingent statements are not used. The data shows that more moderate collusion levels can be sustained when pre-game communication about continuation game actions are allowed but contingent statements are not used. It appears that in these cases more coordination in the continuation game at high payoff levels are sustained if subjects comply with agreements in the first period. However, when agreements are cheated on communication about continuation game actions is not believed even if it would be credible according to the theory. In these cases contingent *behavior* appears to arise that can sustain low levels of collusion but not high levels. We explore in the paper what behavioral theories are consistent with maintaing moderate levels of collusive outcomes in the long run.

Our analysis would fall far short of giving insights into collusion in real markets if the analysis stopped at this point. As soon as one explicitly admits pre-play communication as an essential coordination device in

a dynamic game it becomes impossible to ignore the issue of renegotiation. According to theory, renegotiation should severely limit the ability of players to achieve collusive otucomes in a dynamic game. Players should always renegotiate to a Pareto optimal continuation equilibrium. In our setting the implications are particularly dramatic since in the collusion game we analyze no collusion should be feasible when renegotiation is allowed. Surprisingly, collusion is even more successful when we allow renegotiation by unrestricted chat. The reason is not that the experimental subjects fail to renegotiate to the Pareto optimal continuation equilibrium regardless of first period play. Indeed, subjects mostly comply with this prediction of renegotiation theory. However, despite such non-contiengent play almost perfect collusion is achieved in first period play. We show that this behavior arises because renegotiation does not only have a coordination function for the continuation game, but also generates non-monetary payoffs for the subjects. We show that subjects strongly verbally reproach cheating on agreements during the renegotiation game and that receiving such verbal punishments reduces the likelihood of cheating in later rounds of play with other subjects. In contrast, when renegotiation is restricted to only statements about actions in the continuation game such payoffs arising from social interaction are not available. We show that in these situation the ability to maintain collusion is dramatically undermined. However, there is again no convergence to competitive behavior since as in restricted pre-game communication with contingencies, contingent behavior in the continuation game arises. This again leads to some persistent collusion at a low level in the long run.

The results on renegotiation suggest that social context plays very important and theoretically under appreciated role in collusion. It is clear that the results from this study do not imply that renegotiation will facilitate collusion in real markets, but it suggests that varying the social context in which a collusion game is played is a important future avenue for experimental research into the role of renegotiation in games.

In order to focus on the theoretically central issues in collusion theory, we have developed a design that is quite different from most studies of collusion in the lab. Any collusion game can be thought of as a one period duopoly game followed by a coordination game. Collusion is achieved by making play in the contiunation equilbrium contingent on the outcome of the first period. We capture this structure of collusion games by adopting a two period design. The first period consists has the structure of a three strategy Bertrand game. The second period is a three strategy coordination game with three Pareto ranked equilibria. This design allows us to keep the game cognitively relatively simple, limiting the issue of complexity. The possibile contingencies that could be specified are very limited especially compared to an indefinite horizon game, where in principle an infinite number of contingencies could be specified. Since this is a two stage game we can also let subjects play it many times with different opponents so that we can analyze the long run treatment effect on experienced subjects. Limiting the complexity of the continuation game also allows us to abstract away from issues of bargaining between equilibria in the contiunation game that are not Pafreto comparable, which allows us to focus the analysis on contingencies and abstract from conflicts of interest in bargaining. At the same time this design make0013 T0474otinatioe

2. Collusion and the Theory of Communication in Games

The standard theoretical approach to price collusion is quite simple conceptually: collusion can be supported at some price p^C if the potential gain from undercutting in the short run is less than the long run losses induced by a switch from collusive to competitive behavior in the future. What is critical for the argument is that both the promise of future collusion as a reward for past collusive behavior and the threat of competitive behavior as a punishment for a past deviation are credible in the sense that they involve equilibrium play. A credible threat therefore requires a coordinated switch between to different equilibria of the continuation game. But the multiplicity of equilibria that sustains collusion in repeated games is also the greatest challenge for the theory. Any collusion game involves a coordination problem. How do players coordinate on a specific equilibrium? How can players achieve the knowledge of play that is assumed in the Nash equilibrium concept? In this section we review the impact that communication should have according to theories of cheap talk (see Aumann 1990, Farrell and Rabin 1996) predict about the type of communication we should expect if players are to achieve coordination on a collusive outcome.

In order to better illustrate the theory, we introduce a simple model on which we also base the experiments that follow. Our model exploits the general structure of collusion models just discussed, namely that they can always be described as a one shot oligopoly game followed by a coordination game. Collusion requires that players use the outcome of the one shot game as a coordination device for the play of the continuation game. This is the essence of all theories of collusion based on infinitely repeated games (Abreu 1988, Abreu, Pearce, and Stacchetti, 1990) or finitely repeated games (Benoit and Krishna, 1985).

For the first stage oligopoly game consider a simple symmetric Bertrand duopoly game with three prices: Low (L), Medium (M), and High (H). Let π^i be industry profits if demand is served at price i, and assume $\pi^H > \pi^M > \pi^L > 0$ and $\pi^L > \pi^M/2 > \pi^H/4$. The following matrix (with player 1's strategies being the rows and player 2's strategies the columns) captures the typical payoffs of a one shot homogeneous goods Bertrand game:

$$
(1)\qquad
\begin{array}{cccc}
\textit{Player 1 payoffs} & L & M & H \\
L & \frac{\pi^L}{2} & \pi^L & \pi^L \\
M & 0 & \frac{\pi^M}{2} & \pi^M \\
H & 0 & 0 & \frac{\pi^H}{2}
\end{array}
$$

The unique equilibrium of the game described in (1) is (L,L). In a typical collusion game we would assume that the game is an infinite repetition of the stage game described in (1), with future payoffs discounted by the discount factor δ. Such an infinite horizon game would yield a continuation game with an infinite number of strategies. In order to reduce the strategy space, but still capture the essential features of this infinitely repeated game, we use the following pay-off matrix for the continuation game instead:

$$
(2)\qquad
\begin{array}{cccc}
\textit{player 1 payoffs} & L & M & H \\
L & \Pi^L & \delta[\pi^L + \Pi^L] & \delta[\pi^L + \Pi^L] \\
M & \delta\Pi^L & \Pi^M & \delta[\pi^M + \Pi^L] \\
H & \delta\Pi^L & \delta\Pi^L & \Pi^H
\end{array}
$$

Where $\Pi^i = (\delta/(1-\delta))*(\pi^i/2)$ and δ is the discount factor. Note that with this definition of Π^i, the payoff matrix in (2) has three equilibria, in each of which the players choose the same strategy. These equilbiria

are Pareto ranked with (H,H) the best equilibrium. We refer two the two stage game in which players first play the game in (1) and then the game in (2) as the Two Stage Bertrand Game (TSGB).

One can think of the second stage (2) as the matrix of continuation profits of the infinitely repeated version of (1) when players are restricted to symmetric stationary equilibrium strategies in which players play the same pair of symmetric actions forever. This corresponds to the symmetric optimal punishment equilibria that are often analyzed in applications of collusion theory in industrial organization. Note that in the inifinite horizon version of (1) the optimal punishment is to revert to play of (L,L) forever. Hence, the worst equilibrium in (2) corresponds to the optimal punishment of the infinite horizon game. The payoffs on the diagonal of (2) correspond to the discounted payoffs from the three symmetric optimal punishment equilibria. The off-diagonal payoffs give the discounted payoffs following a deviation in the second period (i.e. the first period of the continuation game). Whenever a player is cheated and has a higher price this gives zero payoffs in the first period and then $\pi^L/2$ for ever. Whenever a player deviates and undercuts price i corresponding to a symmetric equilibrium, the player receives the industry profit π^i in the first period and then $\pi^L/2$ for ever.

We assume that the incentive conditions are satisfied so that {(L,L),(L,L)}, {(M,M),(M,M)}, and {(H,H),(H,H)} are equilibrium outcomes of the TSBG if players play (L,L) in the second stage after any deviation in the first. To abstract from issues of conflicts of interest between players, we will later set the payoffs in such a way that asymmetric actions in the first period cannot be part of an equilibrium strategy for the TSBG. We are now in a position to discuss the predictions theory makes for the type of pre-play communication that we should see for collusion to be sustained in the TSBG.

In order for pre-play communication to have any systematic impact on actual play it must be the case that players take messages to have meaning. But players cannot believe just any message because opponents may have an incentive to lie in order to induce favorable behavior. Aumann (1990) and Farrell and Rabin (1996) have introduced the requirements that m

help to sustain collusion in period 1 given the theory. Note that the same problem would arise if players can only communicate about first period actions and second period actions, but not about second period actions contingent on first period behavior. Suppose the player sends the message: "I will play H in period 1 and in period 2 and I expect you to do the same". The player would want this message to be believed whether he plans to play H in period 1 or not, so it is not self-signaling. It is also not self-committing because if it is believed he has an incentive to undercut. Again, the only messages that are self-signaling and self-committing will involve suggesting (L,L) for period 1.

Messages suggesting cooperation in the TSBG can only be self-committing and self-signaling, *if contingent strategies (i.e. punishment in Period 2 if there is a defection in Period 1) are specified*. Clearly, a message that states a subgame perfect equilibrium strategy is self-signaling and self-enforcing. We should therefore see communication have value in facilitating collusion only if such contingent strategies are used.

One of the problems of this analysis of the effect of communication in collusion games is that it relies purely on pre-game communication. If we believe that communication is important for achieving coordinated outcomes then it is not clear why we should only allow pre-game communication and not communication before every round of the game. The introduction of communication into collusion games therefore immediately begs the question of renegotiation. Suppose in the TSBG the players can communicate before both rounds. Once round two arrives both players are left with game (2). Communication should allow them to obtain the Pareto optimal equilibrium in the game and a suggestion to play this is as we have discussed self-committing and self-signaling. But then there can be no contingent behavior in period 1. It follows that period 1 communication should be conducted ignoring period 2. Messages about period 2 content should always be ignored. Furthermore, the only messages that are self-committing and self-signaling for period 1 will suggest the one shot equilibrium (L, L). No collusion is possible in our model under renegotiation.

The restriction to a continuation game with Pareto ranked equilibria is responsible for the dramatic breakdown of collusion under renegotiation in our model. We know that the scope for collusion under renegotiation is significantly wider under all renegotiation concepts when asymmetric continuation equilibria are allowed (see Benoit and Krishna 1988 for finitely repeated games and Van Damme 1989, Farrell and Makin 1989, Abreu, Pearce, Stacchetti 1987 for super games). For our purposes it is helpful to make the renegotiation problem as stark as possible to get a clean test of the main predictions of renegotiation, namely coordination on the best feasible continuation equilibrium after any history of the game and a limitation on the ability to collude in period 1.

Our discussion of the theory has focused on the standard collusion models that are based on the repeated games literature, assuming that the payoff matrices above represent actual payoffs. However, there may be considerations of altruism, fairness, and spite or presence of conditional cooperators that may have an impact on the outcome. Such considerations may also influence the impact of communication. For example, communication that reveals an opponent's planned action may change the decision of a conditional cooperator. We discuss how to potentially detect such factors in our game when we discuss the treatments.

More importantly, communication itself and the content of communication received (and sent) may have a direct impact on the payoff of the players. We will show that this possibility can make a large difference under some renegotiation settings.

3. Experimental Design

a. The Basic Design of the Experiments

In our view, both collusive strategies and the types of communication that can facilitate collusive behavior require relatively sophisticated reasoning so that we would not expect subjects to spontaneously play equilibrium strategies or be able to distinguish between credible and non-credible communication. For this reason we have chosen a design that allows the subjects to learn strategic interaction as well as the use of communication in the game.

For all treatments, subjects play twenty rounds of the TSBG. A "round" refers to an entire play of the TSBG while a "period" refers to one of the two games played within a single round of the TSBG. Subjects are randomly matched with a new opponent in each round. Sessions are sufficiently large (minimum of twenty subjects) that it is unlikely that there are repeated game effects between rounds.

For the first ten rounds in all treatments subjects play the TSBG without any communication. This allows us to mimic field settings where the players are experienced and presumably fairly sophisticated. While certainly not a perfect substitute for years of experience, ten rounds are sufficient that subjects fully understand the experimental interface, are comfortable with the payoff tables, and grasp the main strategic issues in the TSBG. Subjects' attempts to master these basic issues are not likely to interfere with their ability to effectively use communication. Having ten rounds of play before communication also allows play to converge to the one shot Nash equilibrium in the first period. This makes the task facing subjects more challenging (and more realistic) as they have to overcome a history of non-cooperation rather than merely maintain an existing cooperative agreement.

Treatments vary by the type of communication is available in Rounds 11 – 20. Having 10 rounds of play with communication allows subjects to learn about the use of communication in each of treatments. To the extent that play converges, we can think of play in round 20 as the play that emerges when subjects are experienced at communication.

By using a TSBG our design differs markedly from earlier work on collusion that have used indefinitely repeated games as a proxy for supergames (cite). We moved to the two period game because it has significant methodological advantages. Our goal was to capture in as simple as possible a game all of the most important theoretical properties of collusion games. In particular, we wanted a game in which it was relatively easy to figure out the structure of credible statements. First, we want to make sure that the behavior we observe can be reasonably related to the theory and not be driven simply by the complexity of the collusion problem. The TSBG has the advantage that the number of potential contingencies that an experimental subject has to consider (and potentially communicate about) is relatively limited. As the horizon of the model increases the contingencies that have to be considered increase dramatically. In a game with an indefinite horizon players have to deal with an infinite number of potential contingencies. Our design allows us to abstract from the cognitive problem involved in collusion and focus on the strategic aspects of communication.

Second, in line with our desire to limit complexity we do want to keep the action space relatively small. However, we believe that we limit the conclusions that can be drawn from the experiment if we restrict individuals to only two actions per period as is often done in the experimental literature on collusion (cite). With only two actions the non-colluding action is the default alternative. This makes it difficult to distinguish between failure to cooerdinate and punishment behavior. An advantage of using a game with three actions per period is that subjects have an intermediate choice. Colluding at “Medium” offers some gain over colluding at “Low” with less risk and less need for harsh punishments than colluding at “High”.

Third, we wanted experimental subjects to have sufficient experience with the game to allow behavior to converge. We feel that two hours is the maximum time a session can last without subject fatigue and boredom becoming significant problems. To make sure learning can reasonably converge to long r3915 Tm(me a session0

and 2 makes collusion more difficult to achieve. However, the data reported below indicate that many subjects recognize the relationship between periods. Nevertheless, our approach may systematically underestimate the likelihood of collusion, especially for tacit collusion in the no-communication treatment. There are further restrictions that our setting imposes on the potential patterns of collusion: Tactics such as using asymmetric equilibria in the continuation game to make punishment more palatable, building cooperation gradually by colluding first at intermediate prices, and using less harsh punishments such as tit-for-tat, are viable possibilities in the supergame but not in the TSBG. However, these concerns hold for all treatments, so the comparisons between treatments should remain valid. As a result we see this simplification as a strength rather than a weakness of the design because it allows us to abstract from some of these strategic issues..

b. The Payoff Matrices in the Two Stage Bertrand Game

In picking specific payoffs for the TSBG we had several goals. First, the period 1 stage game had to have a strong Bertrand competition flavor with a unique Nash equilibrium at (L,L). Second, we wanted to avoid a second period game in which subjects would automatically play the Pareto dominant equilibrium to increase the possibility of observing responses to first period play. For that purpose we tried to make (M,M) focal for the period 2 game played in isolation (i.e. when play is *not* contingent on Period 1 outcomes). This makes it natural to think of play moving to either higher or lower prices as a response to the Period 1 outcome. We hope to achieve this by making the equilibrium (M,M) risk dominant (add citation). The following two conditions guarantee that (M,M) is risk dominant relative to (H,H) and (L,L) respectively:

(3) $$\frac{\Pi^M - \delta\Pi^L}{\Pi^H - \delta[\pi^M + \Pi^L]} = \frac{\pi^M - \delta\pi^L}{\pi^H - [2(1-\delta)\pi^M + \pi^L]} > 0$$

and

(4) $$\frac{\Pi^L - \delta\Pi^L}{\Pi^M - \delta[\pi^L + \Pi^L]} = \frac{(1-\delta)\pi^L}{\pi^M - (2-\delta)\pi^L} > 0.$$

Finally, we did not want a threatened switch from the (H,H) equilibrium to the (M,M) equilibrium in period 2 as a response to cheating in period 1 to be sufficient to support collusion in period 1. In other words, we wanted collusion to be attractive but only achievable if subjects are willing to use a strong punishment. Together with our conditions on risk-dominance this feature will help us to distinguish punishment behavior from coordination problems. Since (L,L) in period 2 is both risk dominated and Pareto dominated in period 2, we should not see the use of action L in the long run in period 2 unless it is used for punishment purposes. At the same time players should learn that any punishment other than L cannot be credible.

To achieve the preceding goals, we set $\pi^L = 78$, $\pi^M = 138$, and $\pi^H = 168$. The discount rate is $\delta = 2/3$ and a fixed cost of 24 was subtracted from all payoffs. The two resulting payoff tables are shown in (5). The row is given by a subject's own choice and the column is given by their opponent's choice. Only the subject's own payoffs are shown in the payoff table.

(5)

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

It is easily confirmed that the unique equilibrium in Period 1, played as a one-shot game, is (L,L) and that play of (H,H) in Period 1 can be supported as an equilibrium outcome via reversion to (L,L) rather than (H,H) in Period 2, but not by reversion to (M,M). The payoff from colluding at (H,H) is 25% greater than the payoff from defection and reverting to the bad equilibrium in period 2 (180 vs. 144). Play of (M,M) in Period 1 can also be supported as a collusive equilibrium but is obviously less attractive than play of (H,H). Parameters have been chosen such that (3) and (4) hold – (M,M) is risk dominant in Period 2 relative to (L,L) and (H,H).

c) The Communication Treatments

In this paper we look at four different communication treatments. As a baseline we use a "no-communication" treatment (N-treatment). Second, we use a treatment in which players can only communicate about the pair of first period prices (P1-treatment). This treatment is similar to communications treatments in the previous literature on communication and give us a benchmark for what happens when contingent communication is not feasible. The main two treatments involve communication by chat. On treatment allows only for with pre-play communication (PChat-treatment), the other permits both pre-play communication and renegotiation before second period play (RChat-treatment).

The No Communication Benchmark (N Treatment):

We first run the N treatment as a benchmark for the other communication treatments. The rules for Rounds 11 – 20 are identical to those in Rounds 1 – 10, with no communication allowed. This treatment addresses a potential confound. Sessions with communication have a pause following Round 10 while new instructions are read. Increased collusion following the introduction of communication could therefore reflect a restart effect rather than any direct effect of communication. The N treatment also has a pause following Round 10 for an announcement that the games for Rounds 11 – 20 will use the same rules as are in effect for Rounds 1 – 10. Any restart effect due solely to the pause should be equally present in the N treatment and the treatments with communication, so the N treatment can be used to control for restart effects.

Pre-Play Communication Only about Period 1 (P1 Treatment)
:

The P1 treatment gives subjects the opportunity to send a message prior to the beginning of Period 1 about play in period 1 only. The message space is limited. 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." Each player is shown both parts of both

players' messages at the same time as choices are made for Period 1. The feedback at the end of Period 1 reiterates the messages as well as reporting the outcome for Period 1.

According to the theory, we should expect, that play converges to the one shot Nash equilibrium in the first period. Statements of the type allowed are not self-signaling and therefore not credible. Nevertheless there may be an initial treatment effect either because subjects do not understand this fact and have to learn it or if they believe that others do not understand it.[2] This treatment also allows us to compare our results to the previous literature that has typically only allowed communication about actions but not about contingent strategies (see Davis and Holt 19??).

Pre-Play Chat (PChat Treatment):

Starting in Round 11 – 20, the PChat treatment allows players to communicate using the chat option in version 3.1 of z-tree (Fischbacher, 2007). This is very similar to using an IM program, with continuous back-and-forth communication possible until one of the players makes a decision for the period. There are no limits on how long subjects can chat and no limits on what they can say. Subjects are given no guidance on what they ought to say.

In this treatment we would expect that players would learn to communicate about contingent strategies and establish a high degree of collusion. We therefore hypothesize that in the PChat treatment we should see more period 1 collusion than in the P1 treatment (or the N treatment benchmark).

Renegotiation (R-Chat Treatment)

The renegotiation treatment uses the same communications technology as PChat, but adds the possibility of Period 2 messages. These allow for renegotiation in the spirit of Bernheim, Peleg, and Whinston (1987) or Farrell and Maskin (1989).. Pre-play communication (before period 1) occurs in RChat in exactly the same way as in the PChat treatment. The renegotiation treatment differs from the Pre-play communication treatment only by allowing communication through a chat window after first period actions are observed and before second period actions are chosen.

The theory of renegotiation suggests that regardless of first period behavior subjects should renegotiate to the best continuation equilibrium in period 2. However, then there can be no punishments after defection from a collusive agreement in period 1. Thus, period 1 collusion becomes impossible. We therefore hypothesize that the best equilibrium is achieved in period 2. Furthermore, RChat treatment should have a

---

[2] There are other reasons why first period communication might help. If subjects are conditional cooperators in the sense of Fischbacher, Gaechter, Fehr (1990), then communication should be helpful even if it solely concerns Period 1. Specifically, consider an individual who is willing to cooperate *if he believes his opponent will cooperate*. If his opponent sends a message indicating that both players are expected to cooperate and if this message is believed (i.e. he believes his opponent is more likely cooperate following receipt of such a message), then he becomes more willing to cooperate. With conditional cooperation, prisoners' dilemma games are transformed into coordination games even without repetition. As for other coordination games we expect communication to act as a coordinating device (Blume and Ortmann, 200x).

lower likelihood of period 1 collusion than the PChat treatment. We further hypothesize that, in constrast to PChat, we should see little specification of contingent strategies in the pre-play communication phase because contingencies are irrelevant for actual play if renegotiation is possible.

Note that the RChat treatment also help us to distinguish between some alternative theories of collusion. If contingent behavior is maintained under renegotiation this would indicate motives such a spite playing an important role in sustaining collusion. But there is also the possibility that renegotiation in RChat also has a different role from that considered in theory. For example, communication may affect collusive behavior by generating non-monetary payoffs through expressions of praise or admonishment in the renegotiation phase.

d) Procedures

Sessions were run at Case Western Reserve University in Fall 2006 and Spring 2007. All sessions were run in a computerized laboratory using z-Tree (Fischbacher, 2007). Subjects were recruited from the CWRU undergraduate population via emails. Sessions took between 1½ and 2 hours and average earnings were slightly more than twenty dollars, including a six dollar show-up fee. These payments were sufficient to guarantee a steady supply of subjects. Subjects were paid their total from all twenty rounds of the TSBG. Payoffs were denominated in experimental currency units (ECUs). These were converted to dollars at a rate of 130 ECUs equal $1.

Table 1 summarizes the number of subjects and sessions for each treatment, as well as recapitulating the main features of each treatment. There are three sessions for each treatment with at least twenty subjects per session. We initially tried to implement a "perfect stranger" matching to eliminate any repeated game effects between rounds, but could not get the software to work for this. Even with random matching, the sessions are sufficiently large that any repeated game effects between rounds should be minimal.

[Table 1 about here]

The instructions were read to the subjects, and were also shown on the subjects' computer screens. At several points the payoff tables were projected on an overhead screen for examples, making the payoffs common knowledge. The instructions include multiple examples to insure subjects understand what both players' payoffs will be as a function of their actions. The matching for this experiment is relatively complex (fixed matching for the two periods within a round, random re-matching between rounds), so this point was also emphasized. Following the instructions, subjects were asked to complete a short quiz testing their understanding of the experimental instructions.[3]

The experimental materials are framed using abstract language. For example, subjects are never told they are choosing prices. In Period 1 they choose between "A", "B", and "C" and in Period 2 they choose between "D", "E", and "F", with the three labels in each period corresponding low, medium, and high prices. To ease exposition, the terms "Low", "Medium", and "High" are used throughout this paper even though these are not the labels seen by subjects.

[3] Instructions and slides are available at [add url].

Subjects knew they would be playing a total of twenty rounds of the TSBG. They also knew that the first ten rounds would be played without communication and that there would be a pause after the first ten rounds for additional instructions. The possibility of communication was only introduced at this intermediate point. The instructions for Rounds 11 – 20 in all treatments with communication stress that the rules of the experiment did not change beyond the addition of messages or chat. To maintain parallelism there was a pause prior to the 11th round in the NC treatment with a short announcement that none of the rules would change.

In the sessions with the P1 treatment (limited message space), the instructions prior to round 11 described in detail what messages could be sent (including the option to send "no response") but provided no guidance about why any particular message ought to be sent. The instructions stress that these messages are cheap talk:

> You do not need to use the choice your message says should be made. For example, you can send a message that says you think both you and the other player should choose "A" in Period 1 and then actually choose "B."

Subjects in the two chat treatments received extensive instructions, largely focused on the mechanics of using the chat program. The instructions give the subjects no guidance on what types of messages should be sent other than (i) requesting that they not identify themselves (ii) asking them to avoid offensive language, and (iii) requesting they notify the other player when they made a decision.[4] We did not observe any subjects identifying themselves[David: is this really true??], but there was some use of offensive language following defections in Period 1. The instructions once again stress that the messages are cheap talk with no direct effect on payoffs.

Subjects had printed copies of the payoff tables for *both* periods available whenever they made a decision (there was not sufficient room on the computer screen to include copies of the payoff tables along with the interface for sending messages). When choosing a price for either period, the interface showed subjects any messages or chat from either player for the current period as well as a summary of outcomes for all previous rounds. This summary included both players' prices and payoffs for Periods 1 and 2 for each round, but communication from previous rounds could not be displayed due to space limitations. The interface automatically showed this summary for the three most recent rounds with a scroll bar that could be used to see earlier rounds. When choosing a price in Period 2, subjects could see the prices and payoffs for both players in Period 1, but could not see any communication from Period 1. At the end of each period subjects received a summary of the prices chosen by both players as well as both players' payoffs for the period. Period 2 feedback also included the sum of payoffs across both periods for both players.

[4] A subject's chat window shuts down when he makes a decision for the current period, but his opponent's chat window neither shuts down nor indicates that a move has been made. In pilots this led to subjects trying to send messages to opponents who had already moved and therefore couldn't receive messages. Long delays and great subject frustration resulted. This problem was largely eliminated by having subjects inform the other player when they enter a choice.

The interface (with one exception) did not include identifying information about a subjects' opponent (e.g. an i.d. number) to limit the possibility of repeated game effects across rounds. To make it possible for subjects to tell whether a message had been sent by themselves or their opponent, messages in the chat window were identified with a randomly generated ten digit "chat id". It was not possible to generate new chat ids across rounds. Subjects were not allowed to have any writing implements during the experiment to prevent them from writing down the other players' chat ids, and it seems unlikely that they would have remember long random numbers across multiple rounds. Nothing in the content of the chat indicates that subjects knew when they had played an opponent previously.[David: Again is this really true??]

Sessions were automatically ended at the two hour mark to avoid subject fatigue (this was *not* announced to subjects in advance). Due to this rule, one session of the RChat treatment only had sixteen rounds and another only had eighteen rounds.[5] Subjects were paid privately in cash at the end of the experiment. To limit contamination across sessions, payoff tables were collected from subjects and an announcement was made asking them not to share any information about the rules of the experiment or their experiences in the experiment with other individuals. Informal post-experiment discussions with subjects indicated no evidence of contamination.[6]

4. The Main Results

The purpose of our experiments is to study how the availability and structure of communication affects collusion in the TSBG. Collusive behavior in the TSBG is best captured by Period 1 choices, since setting a price other than "Low" is inconsistent with Nash equilibrium for the Period 1 game (in isolation) and can only be supported in equilibrium by the use of contingent strategies in Period 2. Period 1 choices can naturally be ordered from the least cooperative ("Low") to the most cooperative ("High"). To facilitate exposition of the basic structure of the results we have created a "Period 1 Cooperation" variable. It equals 1 for choice of "High", ½ for choice of "Medium", and 0 for choice of "Low." This variable captures for most purposes the main properties of the period 1 data. In this section we it for our analysis unless a more detailed look gives additional insight.

Period 2 choices primarily help us to detect the degree to which players' decisions are contingent on first period choices and communication. Period 2 choices therefore can help us to distinguish between different possible theories of collusion, communication, and renegotiation. Given this very different role of the second period, we present results on treatment effects only for period 1 choices and analyze period 2 choices only as part of our more detailed analysis of the incentive structure created by contingent behavior in period 2.

a. Behavior in Periods 1 -10

[5] One session of the NC treatment only has nineteen rounds of data due to a software problems that caused Round 20 data to not be saved.

[6] We had no specific reason to expect contamination, but we were unusually careful because CWRU has a relatively small undergraduate population (about 3500 students at the time).

In the initial phase of ten rounds with no communication, Period 1 Cooperation followed a reliable pattern. Modest levels of cooperation in Round 1 (average Period 1 Cooperation = .31) collapsed over time to almost unanimous choice of the non-cooperative action, "L", in Round 10. By round 10, 84% of subjects choose "L". Differences between treatments for Period 1 Cooperation in the first ten rounds are small and not statistically significant. This result suggests that tacit collusion is very hard to sustain in the TSBG in comparison to implementations of indefinitely repeated games (see DalBo 200?). However, we may still detect some potential for tacit collusion by analyzing the degree to which Period 2 behavior is contingent. Period 2 choices are more diffuse than Period 1 choices in Rounds 1 – 10, but by Round 10 a clear mode has emerged at "Medium" with 53% of all choices. Period 2 choices are dependent on Period 1 choices throughout Rounds 1 - 10. Aggregating across Rounds 1 – 10, if the other player's Period 1 choice is "Low", then average Period 2 Cooperation is .51. However, if the other player chose either "Medium" or "High" in Period 1, average Period 2 cooperation increases to .63. This effect of Period 1 play on Period 2 cooperation is statistically significant and strengthens over time.[7] However, this degree of contingent behavior would not be sufficiently strong to prevent "Low" from being the profit maximizing choice for Period 1 against the empirical distribution of actions.[8] This analysis shows that the type of contingent play that can lead to tacit collusion is present in our experiment without communication, but that it is not strong enough to overcome undercutting incentives. Nevertheless, we would expect this effect to push us away from the one shot Nash equilibrium in a model with a continuous action space.

b. Period 1 Short Run and Long Run Treatment Effects

We are interested in the initial and long run responses to the various treatments.[9] The four panels in Figure 1 plot the frequency of each choice in period 1 across Rounds 10 – 20 for each treatment. For the N treatment, there is no sign of a restart effect. Period 1 Cooperation is basically constant between Rounds 10 and 11, and continues a gradual decline through the remaining rounds. In contrast, all communication treatments show a sharp increase in Period 1 cooperation at levels H and M and a sharp decline in the proportion of L played in Round 11 when communication becomes available. But there is a striking contrast between the three communications treatments. In the P1 treatment there is a sharp decline in the use of H (which almost dies out by round 20) and also a steady decline in the use of M. There is no substantial difference in the long run behavior between the P1 and the N treatments.

The dynamic patterns of treatment effects are dramatically different for the chat treatments. In the PChat treatment cooperation at H in Round 11 jumps to almost twice the proportion of the P1 treatment and play of L is almost non-existent (as compared to about 20% in the P1 treatment). After the initial short run

[7] To establish statistical significance, we regressed Period 2 Cooperation on own Period 1 Cooperation, other's Period 1 Cooperation, round dummies, and an interaction term between other's Period 1 Cooperation and the round. Without communication it is not necessary to instrument for either Period 1 choice. Standard errors are corrected for clustering at the subject level. The estimate for other's Period 1 choice is .085 with a standard error of .040. This is statistically significant at the 5% level. The interaction term equals .018 with a standard error of .007, statistically significant at the 1% level.

[8] Assuming the empirically optimal choice in Period 2, the average gain from choosing "Low" rather than "High" for Period 1 is 31.8 ECUs over Rounds 1 – 10. Comparing "Low" with "Medium" yields a gain of 10.5 ECUs. These payoff differences only change slighly over the first ten rounds.

[9] Given the fact that we interpret only the behavior after learning has occurred to correspond to the behavior of experienced individuals we do not find the average treatment effect of particular interest.

treatment effect there is a distinct decline in cooperation at H until Round 16 that is just as sharp as in the P1 treatment.[10] However in contrast to the P1 treatment this decline isfollowed by an increase over Rounds 17 – 20. By the end of the experiment, behavior is indistinguishable from that achieved in Round 11. In Round 20, 72% of the subjects chose "High" and 56% of the pairs successfully colluded at "High". The u-shaped pattern of the use of H in period 1 is statistically significant. [Include here the test we ran aggregating outcomes by two period tranches].

In the RChat treatment, the jump in H is even more pronounced than in the PChat treatment. However, no decline is observedin cooperation at H at all. Instead, cooperation at H remains almost constant at a rate slightly exceeding .8. The small shifts upward in Rounds 17 and 19 should be taken with a grain of salt, as these are driven by RChat sessions ending in different rounds. Breaking the data down by sessions, Period 1 Cooperation is basically constant for Rounds 11 – 20. In Round 16, the last round where all subjects played, 80% of subjects chose "High" and 66% of all pairs coordinated. While the long run effect on the PChat treatment is consistent with theoretical predictions, the observation that renegotiation leads to *higher* levels of Period 1 cooperation contradicts the predictions of renegotiation theory.

[Figure 1 about here]

To confirm the statistical significance of these treatment effects, we fit a linear trend for each treatment, with all the treatments that allow communication differenced from the N treatment. To make it easier to identify treatment effects at the beginning and end points of the experiment's second phase (Rounds 11 – 20), we transform the standard linear model to yield the statistical model:

$$\text{Period 1 Cooperation} = \beta_0\left(\frac{20-Round}{9}\right)+\gamma_0\left(\frac{Round-11}{9}\right)+d_{TR}'\beta_{TR}\left(\frac{20-Round}{9}\right)+d_{TR}'\gamma_{TR}\left(\frac{Round-11}{9}\right)+\varepsilon \qquad (6)$$

The variable d_{TR} is a vector of treatment dummies for the three treatments with communication (P1, PChat, RChat). The model transformation allows us to interpret the β parameters as the short run treatment effects and the γ parameters as the long run treatment effect. The parameters β_o and γ_0 therefore estimate, respectively, the starting and ending levels of Period 1 Cooperation for the NC treatment. The vectors of parameters β_{TR} and γ_{TR} estimate, for each of the other five treatments, *differences* from the N treatment for starting and ending levels of Period 1 Cooperation. The error term is given by ε and satisfies the usual assumptions.

Table 2 presents results from two OLS regressions with the standard errors corrected for clustering at the subject level, where Period 1 Cooperation is the dependent variable[11] [KUK: Here we should present

---

[10] However, in constrast to the P1 treatment, the reduction in the choice of H is accompanied by an increase in the choice of M.

[11] A natural concern with our regression analysis is that the use of OLS models with standard errors corrected for clustering at the subject level might cause us to misrepresent the data. Either a session effect or the discrete nature

[redacted]

The starting point estimates (differenced from the N treatment) in all communication treatments are large and statistically significant Consistent with our observations from Figure 1, the difference between the estimated start point of each of the two chat treatments and the P1 treatment is significant at the 1% level. The difference between the start points for the two chat treatments is also significant, but only at the 5% level.

The end point estimate for the P1 treatment is relatively small, but still statistically significant at the 5% level.[13] Looking at Figure 1, this reflects the small but fairly constant difference observed between the P1 [redacted] difference is not significant at even the 10% level for the PChat treatment. However, there is a difference between the PChat and RChat treatments that is statistically significant at the 1% level.

Conclusion 1: Both chat treatments yield a large improvement in Period 1 Cooperation both in the short and in the long run.

Conclusion 2: As predicted by theory, there is effectively no long run treatment effect for the P1 treatment

---

of the depende~~nt variable~~ could cause us to misestimate statistical significance. As a robustness check, Models 1 and 2 have bee

Conclusion 3: The renegotiation treatment leadsto strictly higher Period 1 collusion than the PChat treatments. We therefore strongly reject the predictions of renegotiation theory.

c. Verbal Agreements and Cheating

In all three communications treatments we can identify agreements between the player to play some kind of action in period 1. We call it angreement for the P1 treatment if the two players make coinciding suggestions for play. In the PChat and RChat treatments we essentially use the same definition, although our coding of the chat allows us to make some finer distinctions as we discuss in section 5. However, these do not materially affect the results. Lack of agreement is rare even in the P1 treatment where players are restricted to only one round of simultaneous suggestions of play. [Include percentages]

Figure 2 shows the three panels that plot the percentage of agreements on H, M, and L over time for the three communication treatments. Note that in all three treatments players overwhelmingly tend to agree on both playing H in period 1. In the P1 treatment we initially see 70 to 80 percent of agreements to be on H. This somewhat declines later on, as H play deteriorates, but goes back to about 70% as M play deteriorates as well. Note that there is an enormous difference between the agreement and the actual play, which implies that there is a very high probability of cheating on agreements. In accordance with cheap talk theory, communication is clearly not credible in this treatment. However, it remains rational to still make an announcement of H as long as there is a small probability that some opponent takes the agreement at face value. This generates a large incentive to state "H" but never to comply with it.

[Insert Figure 2]

In PChat the proportion of agreement on H is between 80 and 90 percent. Given the actual behavior documented in Figure 1 this still implies a very significant rate of cheating on agreements. However, communication becomes more credible at the end. The interesting question that arises from this observation is whether players learn to use language in such a way that agreements eventually become more credible in PChat. In particular, the theory of cheap talk would lead to the predictions that agreements that include understandings of history dependent contingent play in period 2 should be more credible than non-contingent agreements. We explore this issue further in the next subsection.

In RChat the rate of agreement on H is even higher and between 90 and 100 percent. There is still some cheating on the agreement, but it is notably lower than in the PChat treatment. The relatively high degree of credibility of first period agreements in RChat is again a puzzle given renegotiation theory.

d. The Role of Explicit and Implicit Contingencies

Our leading contender for explaining period 1 collusion at H is contingent behavior in period 2. Only with period 2 contingent behavior can there be any incentives to comply with a collusive strategy according to all theories of dynamic collusion we have. Furthermore, according to the theory of cheap talk it would be

crucial for players to communicate that they will engage

play of H in period 2 after cheating in period 1 is about 55%. However, this steadily declines over time and reaches only 30% by period 17. Hence, the degree of conditioning on period 1 play gets larger and larger over time. If high degrees of cooperation in period 2 lead to cheating in period 1, this observation is consistent with cooperation initially deteriorating in PChat and then recovering as conditioning on first period play gets more pronounced. In contrast, play of H in the renegotiation game never falls below 60% and is often close to 80%. Overwhelmingly, subjects renegotiate to the best outcome after cheating in period 1. In fact, degree of contingency becomes smaller over time. However, there is no response in terms of a deterioration in period 1 collusive behavior. Hence, there is a clear period 2 renegotiation effect as predicted by renegotiation theory, which makes period 1 behavior even more puzzling. We seek to resolve this issue through our content analysis in section 5 of the paper.

While we have made a clear observation of increased contingent behavior in period 2, it appears critical from the theory that contingent play is agreed to explicitly, to make collusive behavior occur in the PChat treatment. In contrast, we should see little contingent agreements in the RChat treatment if its is anticipated that renegotiation would occur. In the analysis of Chat that we specify in more detail below we have distinguished between non-specific threats to punish deviators from a period 1 agreement and specific threats to play L in period 2 if any deviation occurs in period 1. Figure 5 shows that there are very strong differences between the two treatments in the use of such threats. In the renegotiation treatment players use threats much less frequently and they use almost exclusively non-specific threats. Furthermore, in the PChat treatment, the great majority of threats is specific to play L in response to cheating in period 1.

[Insert Figure 5]

Note that the use of specific threats is increasing over time in the PChat treatment, suggesting that the increasing use of explicit contingencies is a leading candidate for explaining the turnaround in period 1 play of H in this treatment. However, it should be noted that the use of contingencies remains fairly infrequent. To understand more deeply whether the use of contingencies or other aspects of communication drive the learning behavior in PChat and also what may cause the counterintuitive behavior in RChat, we now turn to a more detailed analysis of the content of chat in the two chat treatments.

5. Content Analysis of the Chat Treatments

At this point we reach the question that is at the heart of our paper. What is it about the use of chat that leads to such a dramatic improvement in Period 1 Cooperation? Is it indeed the dramatic success of the contingent strategies we demonstrated above or are there other aspects of unstructured communication that have an important influence on period 1 communication. Or is it simply the fact that we have made it much easier for an agreement to be reached? To sort out these issues we now analyze in detail the messages that are sent and establish what messages do the best in fostering collusion in the TSBG.

a. The Approach to Content Analysis

We developed and implemented a systematic scheme for coding message content. The goal was to quantify any communication that might be relevant for the play of the game, avoiding prejudgments about which sorts of messages were important and which were not. Our methods paralleled those employed by Cooper and Kagel (2004) and Brandts and Cooper (2007). We began by randomly selecting a test sample. This consisted of two randomly chosen dialogues (i.e. the chat between a pair of subjects) for each period of each session with chat. Both co-authors independently developed coding schemes for the test sample. In a series of meetings we reconciled these individual efforts into a single coding scheme.

Two research assistants, one at Michigan and one at FSU were then trained to do the coding and independently coded all messages. Some modifications to the coding scheme were made in response to feedback from the coders.[14] No effort was made to force agreement among coders – the goal was to have two independent readings of each message so that any coding errors were uncorrelated. At no point in the coding process was either RA informed about any hypotheses the co-authors had about the messages. The RAs were repeatedly and explicitly told that their job was to capture what had been said rather than why it was said or what effect it had. Coding was binary – a message was coded as a 1 if it was deemed to contain the relevant category of content and zero otherwise. We had no requirement on the number of codings for a message – a coder could check as many or few categories as he deemed appropriate. A number of the categories have sub-categories. For example, Category 6 ("Explanation") codes the explanation subjects gave for a proposed course of action. There are ten sub-categories under this, four of which code what specific proposal the explanation refers to and another six which code the nature of the explanation. A coder was free to check as many or few sub-categories as they desired when the corresponding category was checked off. For example, consider the following message (all samples from subject dialogues are quoted verbatim, without any corrections to grammar or spelling), "b because theres less risk of lyers, and F because we both benefit most." This was coded by at least one coder under Categories 6a, 6e, 6g, and 6h.[15] Our analysis of the codings uses averages across coders unless otherwise noted. We implicitly assume that errors are independent across coders so that the total error is reduced by averaging. Cross-coder correlations for major categories averaged .66, about the same as earlier studies using the same methods.

A disadvantage of using such a detailed coding scheme is that it becomes difficult to see the forest for all the trees. To make analysis of the codings more manageable, we restrict our attention to the most common categories. Specifically, this was any category observed for at least 7.5% of all relevant dialogues.[16] We initially set the threshold at 10% and then checked whether lowering the threshold to

---

[14] Specifically, we originally planned on classifying messages which agreed with a proposal by the strength of the agreement. This was abandoned when it became clear that the strength of agreement was too ambiguous to be coded accurately.

[15] This message was not coded under Category 1 because it referred to an existing proposal of actions rather than making a new proposal of actions. The earlier message where the actions B and F are proposed is coded under Category 1.

[16] We define a category as being observed for a dialogue if it was coded at some point during the dialogue. The "observe" variable is set equal to ½ if only one coder coded the category and equals 1 if both coders coded the category. "Relevant" dialogues are those where the category in question could have been coded. Thus, for categories relating to Period 2 messages, only Period 2 dialogues in the Chat-ALL treatment were relevant.

7.5% would lead to inclusion of any variables that had a statistically significant effect on outcomes. As will be seen, several of the added variables had a significant impact. We then checked the effect of lowering the threshold to 5% and found that none of the newly added variables had a statistically significant impact on outcomes.

Table 3 summarizes the frequency and effect of the most common categories. Data on Period 1 messages has been combined for the Chat-1 and Chat-ALL treatments. Before going into the data, some additional discussion of the common categories, with a focus on why they might impact collusion, is in order. The categories that propose a course of action (1b, 1c, 1f, 11c) or (dis)agree with an existing proposal (2a, 2b, 12b) are self-explanatory.

[Table 3 about here]

Category 3c is used for threats that explicitly call for use of "Low" as a punishment.[17] Other specific threats were almost never coded, although non-specific threats are almost as common as specific threats of "Low" (6.1% of observations).[18] The use of a specific threat is important, since not all punishments are sufficient to support collusion at "High" in Period 1.

Category 6e is used for messages that point out that collusion will increase both players' payoffs.[19] Brandts and Cooper (2007) find that messages that point to the joint gains from coordination lead to significantly higher coordination in minimum games. Since supergames at their core are a complex form of coordination game, there is some reason to believe this result might extend to the current case. Any message that referred to the payoff tables was coded under Category 6i.[20] Most of these messages reference specific numbers drawn from the payoff tables, although we also included messages that pointed to the payoff table in a more general way.[21] Given that determining incentive compatibility requires making explicit calculations based on the payoff table, it is striking that relatively few dialogues actually refer to the payoffs.

Categories 8a, 8c, and 8e all relate to trust. The first, Category 8a, was coded for messages where a subject indicated that he should be trusted. We did not require that the word "trust" be used explicitly, but instead left to the coders' judgment whether a request for trust was implied. The most common wording was to "promise" to take a certain action, with the implication that promising made the claim more credible.[22] As noted previously, the largest benefit of choosing "Medium" rather than "High" was insurance against the other player choosing "Medium". If promises are believed, the desire to choose "Medium" for insurance should be lessened. Category 8c was coded for messages where a subject indicated a lack of trust in the other player.[23] Once again, we did not require that the word "trust" be used explicitly (although often it was). Finally, Category 8e was coded for messages that request trustworthy

[17] e.g. "i'm choosing d if you don't choose c"
[18] e.g. "well ill just screw u in periodd 2 if u screw me in round 1"
[19] e.g. "if either one screws the other over we both make less"
[20] e.g. "just pick c and f and we will get 180 each"
[21] e.g. "You have to make sure to check the table at the bottom [the Period 2 payoff table]"
[22] e.g. "ok, C and F PROMISE"
[23] e.g. "how can i trust u?"

behavior from the other player.[24] Such messages ty

reflect a general conversation of trust. Suppose Category 8a (indicating you should be trusted) has a positive effect on cooperation while Category 8c (indicating a lack of trust) has a negative effect. If the two types of comments tend to go together, raw statistics like those reported on Table 8 will understate the size of each category's effect on Period 1 cooperation. Even if we control for the time period, treatment, and other categories coded, identifying a causal relationship between the chat content and Period 1 cooperation remains problematic. Comments from some particular category could correlate with Period 1 cooperation because they cause cooperation, but these could also correlate with Period 1 cooperation because individuals who are likely to cooperate are also likely to make certain types of comments. In other words, there may be a confound between the effects of chat content and the individual effects in the data. Finally, any attempt to study the effects of chat is affected by "reflection".[28] Any conversation between two players is inherently interactive. It is tempting to interpret receiving some particular comment as causing cooperative action, but the comment's sender may be responding to earlier messages from the recipient. Controlling for categories that the recipient has coded for previously does not solve the problem, since these comments can in turn reflect even earlier statements.

The regressions reported in Table 4 are designed with these issues in mind. These are OLS models. Robust standard errors are reported in parentheses. An observation is a single play of the TSBG from the Chat-1 and Chat-ALL treatments. Because of the reflection problem, we consider the two choices in a game as a single observation rather than considering the two players' choices separately. Specifically, the dependent variable is "Group Cooperation, defined as the average Period 1 Cooperation across the two subjects playing a game.

[Table 4 here]

The independent variables of interest measure whether each of the most comment categories was observed in the dialogue prior to making decisions for Period 1. This is an average across the two coders and therefore takes on values of 0, ½, and 1. All of the regressions include round dummies to control for effects due to when categories were likely to be coded. The parameter estimates for the round dummies are not of direct interest and are supressed on Table 9 to save space. Model 1 includes a dummy for the Chat-ALL treatment. Equation 7 shows the specification being estimated in Models 2 – 4. These models include fixed effects for the two subjects that are playing the game. Specifically, a parameter is estimated for each subject in the population and this vector of parameters is multiplied by a vector of dummies which equals 1 for the two subjects in the pair and zero otherwise. The dummy for the RChat treatment must be dropped to avoid collinearity. With the inclusion of fixed effects, the estimated parameters for the coding categories capture the change in Group Cooperation from what would be expected *given the two subjects being matched*. In other words we are controlling for any confound between chat content and the individual effects in the data.

(7) Average Period 1 Cooperation = $\beta'_{Round}\, d_{Round} + \beta'_{Category}\, Observed_{Category} + \beta'_{Subject}\, d_{Subjects} + \varepsilon$

[28] This is a well-known problem from the empirical literature on social interaction effects. To quote Manski (1993), "The problem is similar to that of interpreting the almost simultaneous movements of a person and his reflection in a mirror. Does the mirror image cause the person's movements or reflect them?"

Model 1 and Model 2 differ only in whether subject fixed effects are included. While the results are similar, inclusion of fixed effects does have some effect. Specifically, without the inclusion of fixed effects it appears that Category 2b has a significant positive effect on Group Cooperation. When fixed effects are included in Model 2 the size of the parameter estimate is more than halved and statistical significance vanishes. This illustrates the importance of controlling for individual effects.

Looking at Model 2, the parameter estimates for Categories 1b and 1c are both significant at the 1% level. This result confirms our observation from Table 8 that Period 1 cooperation responds strongly to whether cooperation at "High" was suggested as opposed to cooperation at "Medium". Neither disagreeing nor agreeing to a proposal (Categories 2a and 2b) have a significant effect on Group Cooperation. This runs contrary to the impression given by the raw data in Table 3. There is little evidence to suggest that improved Period 1 cooperation with open chat is driven by an improved ability to coordinate strategies. The parameter estimate for Category 3c is positive and significant at the 1% level. The raw data in Table 3 suggested that the effect of Category 3c was modest, but the effect of threatening to punish non-cooperation with "Low" in Period 2 is actually quite strong. Finally, Category 8a has a modest but statistically significant effect on Group Cooperation. Indicating you should be trusted, often by promising to choose "High" in Period 1, does have a positive impact.

Models 3 and 4 break down the data by treatment. Two notable differences emerge. The effect of Category 3c is stronger and more statistically significant in the PChat treatment. This seems sensible, as statements about Period 2 actions are likely to seem less credible when renegotiation is possible. Category 6e, coded for appeals to joint payoffs, has almost no effect in the PChat treatment but has a positive and statistically significant effect in the RChat treatment. There is no obvious reason why such comments should matter more when renegotiation is available.

We have done a number of specification checks to verify the results shown on Table 4. Specifically, Model 2 has been rerun using an ordered probit rather than an OLS model,[29] using perfect cooperation (a dummy for both players choosing "High") as the dependent variable rather than average Period 1 cooperation, and using an IV regression where the chat categories are instrumented with the lagged categories sent by each player as well as the players' lagged choices.[30] The results are not identical across these various specifications, but the main results described above are fairly robust. Most notably, the parameter estimate for Category 3c is always positive and statistically significant. The positive effect on cooperation of specifying an incentive compatible punishment is the single most robust finding from the chat treatments. The signs of the parameter estimates for the other categories that are statistically significant in Model 2 are the same in all three of our specification checks, and Categories 1b, 1c, and 8a are all statistically signficant in two of the three specification checks.[31]

---

[29] Use of an ordered probit is problematic since about a fifth of the observations are completely determined due to the inclusion of fixed effects.

[30] We were concerned that there might be a common uncontrolled factor that drove both the codings and the decisions whether or not to collude. Estimates from the IV regression were not very precise, so we have focused on the OLS results.

[31] Category 1b is not significant in the regression using perfect cooperation as the dependent variable and Categories 1c and 8a are not significant in the IV regression. The parameter estimate for Category 2a is negative and statistically significant in the ordered probit regression. The parameter estimate for Category 2b is negative and

As an alternative to the specification used for Table 4, we have also run regressions where the dependent variable was the Period 1 Cooperation by a single subject in a single game. This regression has two observations for each play of the game rather than one, so standard errors are corrected for clustering. The primary independent variables are whether the subject *received* a comment coded for the category in question prior to Period 1 play. The independent variables are otherwise the same as in Model 2 in Table 9. This specification suffers from reflection problems, as described above, but we nonetheless feel it is useful for getting a handle on whether the effects observed in Table 9 are due to the effect of receiving rather than sending a particular message. Even taking the results with a grain of salt, there are interesting contrasts between this new regression and Model 2. While the effect of Category 1b is largely unaffected, the magnitude of the estimate for Category 1c is dramatically smaller (.044 vs. .237 in Model 2) and statistical significance is somewhat weakened ($t = 1.96$; $p \approx .05$). This suggests that subjects who suggest use of "High" tend to follow this suggestion, but that the effect of recipients is small. The parameter estimate and statistical significance for Category 3c are almost unchanged. The impact of Category 3c is the single most robust feature of our content analysis, and it seems to be equally important to send and receive such messages. The estimate for Category 8a is also little affected. This suggests the effect of making promises is not solely that the promiser feels a moral obligation to follow through, but rather that the recipients of promises tend to believe them and are therefore more willing to take the risk of behaving cooperatively. Finally, the parameter estimate for Category 8e more than doubles (.073 vs. .034 in Model 2) and achieves statistical significance at the 5% level ($t = 2.25$; $p < .05$). Calling for trustworthy behavior doesn't make the sender any more cooperative, but it does have an effect on the receiver.

b. Content Analysis of PChat

The puzzling feature of the communication data on PChat is that the only two communication categories that matter (threats, and statement that one can be trusted) are among the categories least frequently used. But as we have seen subjects use these strategies more frequently over time. In this section we analyze whether behaviors that are associated with successful first period collusion are learned over time.

Contingent Agreements

At first blush, the ,low use of contingent threats suggests that subjects fail to grasp the value of threatening punishment, but a more accurate reading of the data is that the role of explicit (and implicit) contingent behavior is successfully learned, albeit slowly. In Rounds 11 – 15, only 5% of the dialogues were coded for Category 3c by at least one of the coders. This figure rises to 15% of the dialogues for Rounds 16 – 20. Subjects take a long time to try threatening punishment, but once a subject was coded for sending a message in Category 3c, they tended to keep sending such messages. Comparing subjects who were and were not coded (by at least one coder) for sending a message in Category 3c in round t, the relative probabilities of being coded for sending a message in Category 3c in round t + 1 (i.e. the following round) were 63% versus 3%. Threatening to punish is a fairly subtle strategy. Not only does it require subjects to figure out that threatening punishment makes collusion incentive compatible, it also require subjects to reason from the other player's point of view. Cooper and Kagel (2005) present

---

significant for the IV regression. The sign of this estimate is opposite what we would expect, reinforcing our lack of faith in the IV results.

evidence that this is a particularly difficult concept to master, but one that is critical to the development of strategic play. As such, it is not surprising that subjects do not often spontaneously use threats of punishment.

Instead, the data suggests that subjects often learn to use Category 3c by observing others using such messages. Consider subjects who were not coded for *sending* a message in Category 3c in round t. Consider subjects in this subset of the data who did not *receive* a message coded for Category 3c in round t with those who did. The probability of being coded for *sending* a message in Category 3c in round t + 1 is 2% versus 10%. Receiving a threat of punishment makes subjects who have not used such threats five times more likely to try this approach in the next round.

However, the use of contingent agreements gives only part of the explanation why collusive play at H recovers in the last rounds of the PChat treatment. This is clearly illustrated by Figure 6. It plots the proportion of subjects that deviate from a non-contingent agreement on H. This dramatically increases until round 17 and then dramatically declines.

[Insert Figure 6]

Why would there be a reduction in deviations from non-contingent agreements when contingent agreements become more frequent? Again the data suggests that this is generated from a learning effect: Subjects that have been exposed to a contingent threat in the past are not only more likely to suggest a contingent threat. They are also less likely to deviate from a non-contingent agreement (see Figure 7).

[Insert Figure 7]

The data therefore suggests that the effect of explicit contingencies can both increase the explicit use of contingencies but also increase the implicit consideration of contingencies. This factor seems to be an important one in reducing the incidence of cheating on collusive agreements. Overall, the data presents a clear picture. It is hard to learn the strategy of threatening to punish with "Low" in Period 2, but subjects do gradually learn this approach through a combination of insight and, more often, imitation. Once they learn this approach, it works well and they tend to stick with it. Where they do not use contingent statements they nevertheless behave as if they are more concerned about the potential contingency.

While all of these features are very supportive of the reasoning in the standard theory of collusion, we have not shown that the credibility of punishments that is important to this approach matters in our experiments. Notably, a punishment threat is only credible when it is understood to be mutual. Whether this is the case or not is not clear from the chat data. Hence, there is no clear data whether the subject that makes a threat also expects to be subject to this threat if it deviates itself. Furthermore, it is very rare that the proposer of a contingent agreement deviates. However, in the five instances we observe such deviations the other player always carried out the punishment action L.

Trust

As a point of comparison, consider learning to use Category 8a, messages indicating you should be trusted (largely promises). This type of message is used slightly more frequently than Category 3c and,

like Category 3c, has a positive and significant effect on Period 1 cooperation with the effect being roughly equal for senders and receivers. In spite of these apparent similarities, learning is not as strong for use of Category 8a. In Rounds 11 – 15, 11% of the dialogues were coded for Category 8a by at least one of the coders as compared with 15% of the dialogues for Rounds 16 – 20. As with Category 3c, subjects who receive a message from Category 8a are more likely to send such a message in the next round. Consider subjects who were *not* coded for sending a message in Category 8a in round t. If they did not receive such a message in round t, there was 6% of being coded for sending a message in Category 3c in round t + 1 as compared with a 16% chance if they did receive a message from Category 8a in round t. The reason that growth of Category 8a is relatively weak compared to Category 3c is that subjects who use Category 8a do not have as strong a tendency to keep using it in future rounds. A subject who was *not* coded (by at least one coder) for sending a message in Category 8a in round t has a 7% probability of being coded for sending a message in Category 8a in round t + 1 versus a 19% chance of being coded in round t + 1 for sending this type of message if a subject was coded for sending a message in Category 8a in round t. Category 8a is initially far more common than Category 3c, but this difference fades over time. The reason underlying this is that, *subject to being tried*, subjects are more likely to stick with Category 3c than Category 8a.

[Expand on effect on deviating from a non-contingent agreement]

This leads us to a pair of conjectures. In trying to determine what messages are important for generating collusion, taking a dynamic view of the data is essential. Threatening to punish non-cooperation with use of “Low” in Period 2, Category 3c, is an effective strategy that gains popularity with experience. Between the observed learning and the likely effects of selection, we conjecture that our data understates the role that threats of punishment are likely to play in field settings. There has been a focus in the psychology (e.g. Kerr and Kaufman-Gilliland, 1994) and economics (Charness and Dufwenberg, 2006) literatures on the value of promises. Our data certainly indicates that promises play a positive role in promoting cooperation, but the relative growth rates are in favor of punishment. We therefore conjecture that the threat of punishment that underpins game theoretic models of collusion is more important in field settings than the use of promises.

c. Content Analysis for RChat

As we have seen, first period chat seems to give very few clues why RChat leads top so much more collusive outcomes than PChat. The central difference between the two treatments is the presence of a chat opportunity prior to the period 2 decision after period 1 decisions have been observed. We therefore focus in this sec tion on the content analysis of period 2 chat.

Coding categories in period 2 chat closely follow those of period 1 chat. Categories 12b and 15c require no explanation as they correspond to Categories 2b and 6e for Period 1 messages. Categories 10a, 10c, 10e, and 10g code for responses to the first period outcome. Category 10a is for messages giving positive feedback after the subjects have successfully colluded in Period 1.[32] More interesting are the messages associated with non-cooperation. Going out of order, Category 10g was coded for messages where the subject admonished the other player for cheating. These reactions were very frequent in cases where

[32] e.g. “Alright. Fantastic playing with you!”

cooperation was not achieved – Category 10g was coded by at least one coder for 52% of observations from the RChat treatment where the subjects did not collude at "High" in Period 1. The messages were often quite intense, quite personal in nature, and frequently ignored the instructions to avoid cursing.[33] These messages are best interpreted as strong non-pecuniary punishments in the spirit of Fehr and Gachter. As such, we would expect the frequent use of punishments to be associated with greater Period 1 collusion. Frequently coded messages from subjects who cheated either apologize for their actions (Category 10c)[34] or try to rationalize their actions (Category 10e). Some of these explanations may be legitimate, with subjects stating that they didn't choose "High" because they thought the other player wouldn't choose "High".[35] Another common justification was to refer to past cheating by other opponents.[36] This wasn't necessarily truthful – one memorable subject choose "Medium" for Rounds 12 – 20 and justified it every time by saying their opponent had done the same to them in the previous round. In fact, all of their opponents is Rounds 11 – 20 choose "High" in Period 1! The overall dynamic is clear. Those who cheat try to muddy the water in order to avoid punishment in Period 2 while those who have been cheated on seize the costless punishment afforded by the chat. This observation suggests an explanation for why the prediction of renegotiation theory fails in the RChat treatment. Subjects invariably *do suggest coordination at "High"* even when cooperation failed in Period 1, consistent with renegotiation. However, the opportunity to communicate after Period 1 gives an opportunity to punish cheaply. As we establish below, this plays an important role in the strong Period 1 cooperation achieved in the RCha treatment.

Table 10summarizes the data about the use and effect of Period 2 messages. A couple of features are worth noting. First, attempted renegotiation is quite common. Following non-cooperation in Period 1, 91% of the dialogues include a suggestion that "High" should be played in Period 2. These attempts seem to be successful, as choice of "High" is substantially more likely following non-cooperation when Category 11c is coded (see the final two columns on Table 8). Likewise, appealing to joint payoffs (Category 15c) has a positive effect on Period 2 cooperation. Thus, the logic of renegotiation seems to be sound – subjects often negotiate away from harmful punishments.[37] To understand why allowing renegotiation does not lead to lower cooperation, consider the effect of Category 10g, coded for messages admonishing a player for cheating in Period 1. Table 6 considers Period 1 cooperation in round t by subjects who cheated (chose something other than "High") in Period 1 of round t – 1 (i.e. the previous round). The data is broken down by what sort of punishments (if any) were used by their opponents in round t – 1. These punishments can take the form of either using a price other than "High" for Period 2 or admonishing the individual975 Tm(– 1 (i.e. tb)Tj 10.98 0 0 Tmer forh(eati.wn)Tj 10.98 0 0 10.98 453403729 229.79919 T

punishment with a lower Period 2 price.[38] Thus, allowing for communication prior to Period 2 makes a very powerful form of punishment available to players. The threat of being admonished is sufficiently daunting that it overcomes that detrimental effects of renegotiation and leads to higher rates of Period 1 cooperation in the Chat-ALL treatment.

[Table 6 about here]

Conclusion 5: High levels of Period 1 cooperation in the Chat-ALL treatment can be attributed to the potency of verbally punishing individuals who cheat in Period 1

[38] Subjects appear to view the two forms of punishment as complements, as use of admonishment is positively correlated with chosing a Period 2 price other than "High". This runs contrary to the results of Xiao and Houser (2005).

Treatment Effects: Period 1

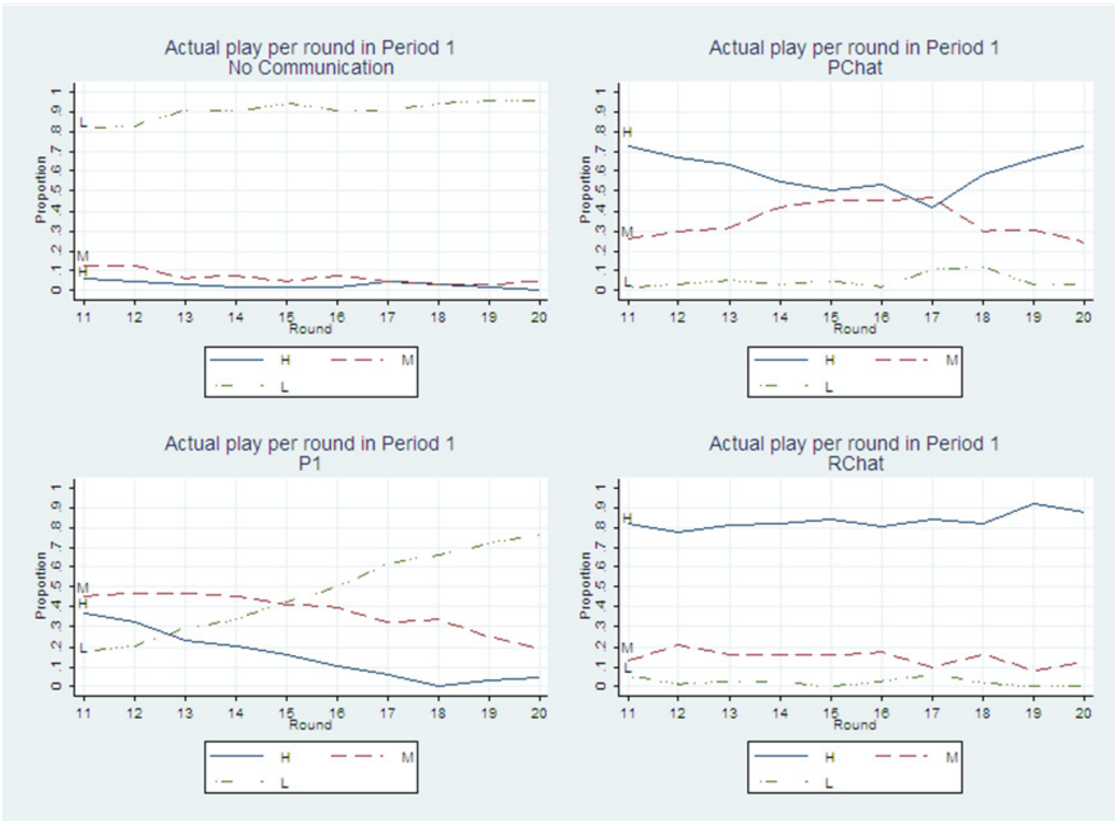

Figure 1

Agreements for Period 1 Play

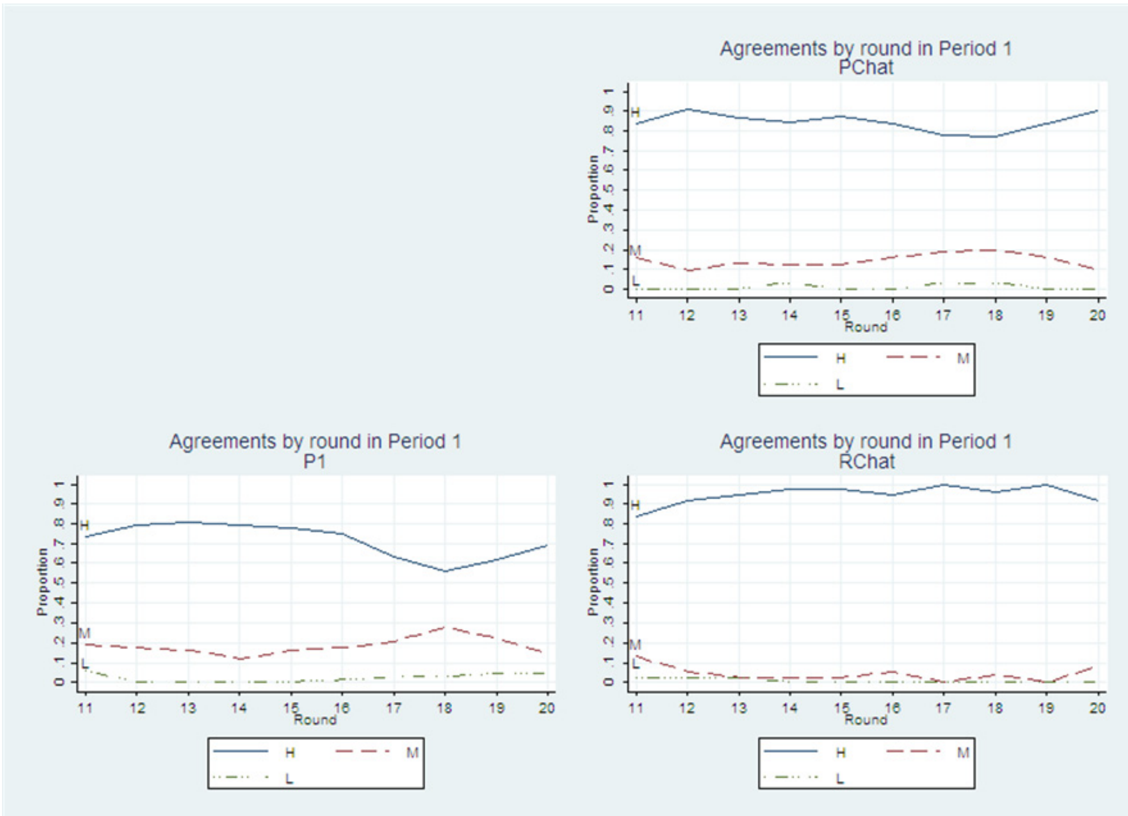

Figure 2

Treatment Effects Period 2

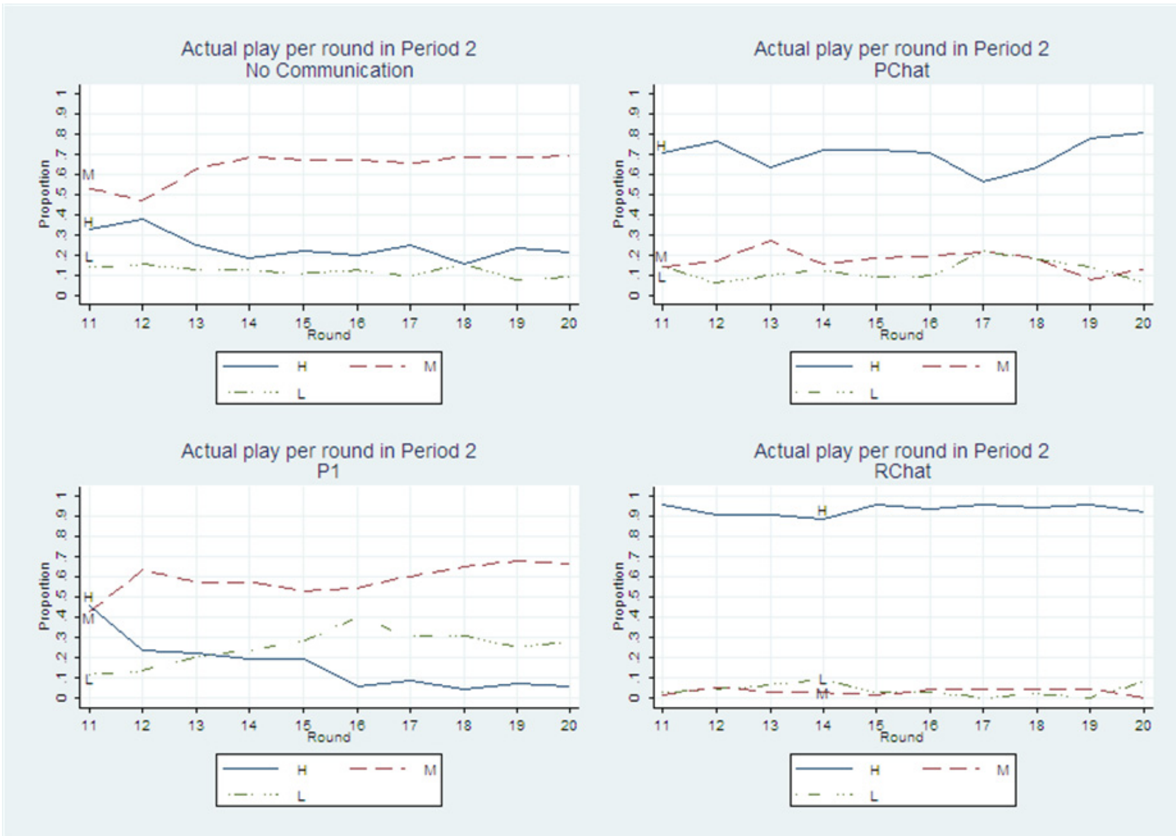

Figure 3

Differences in Contingent Behavior

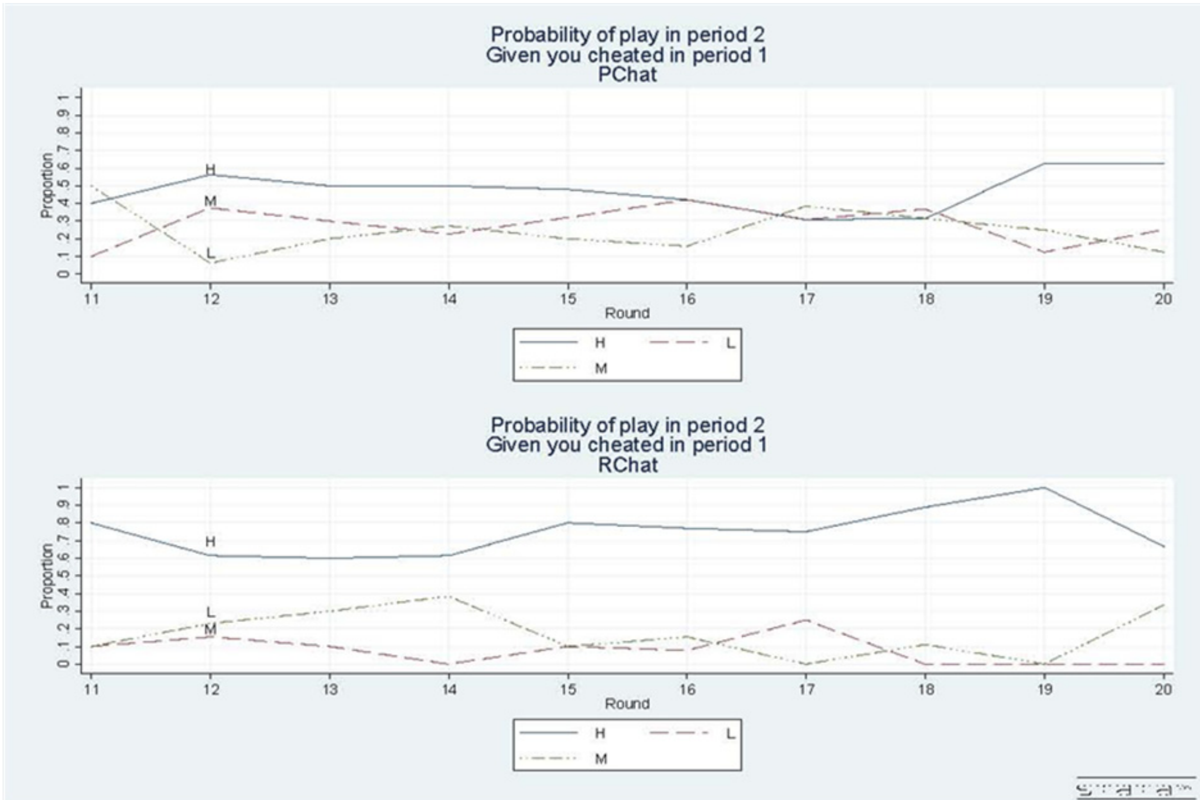

Figure 4

The Role of Threats

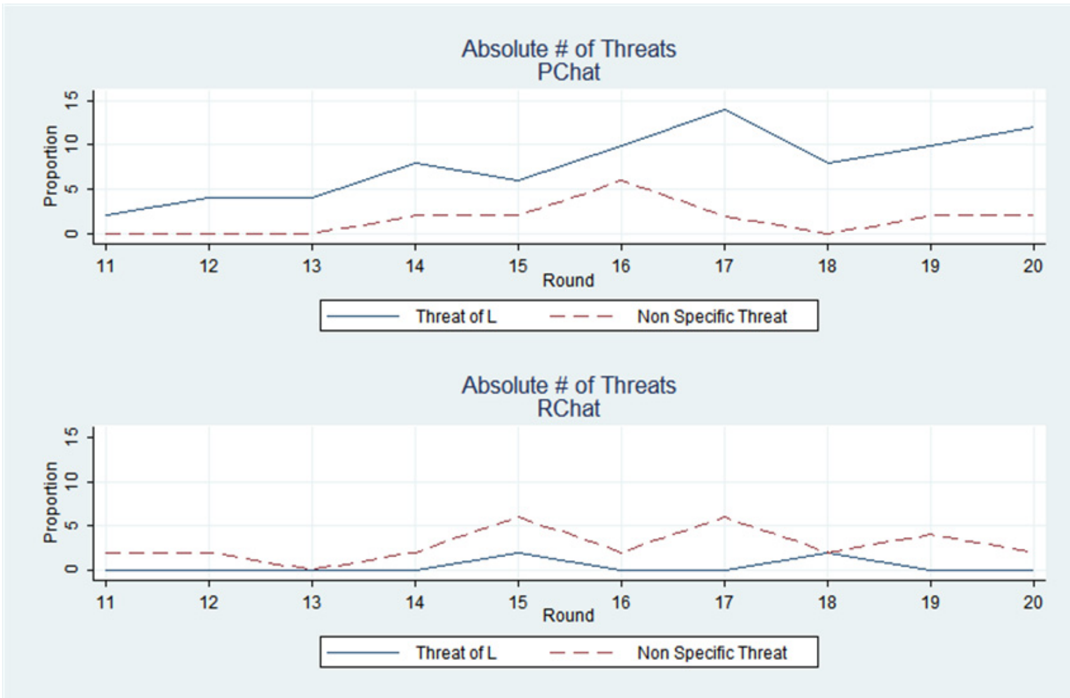

Figure 5

The Turnaround in Cheating on Non-contingent Agreements

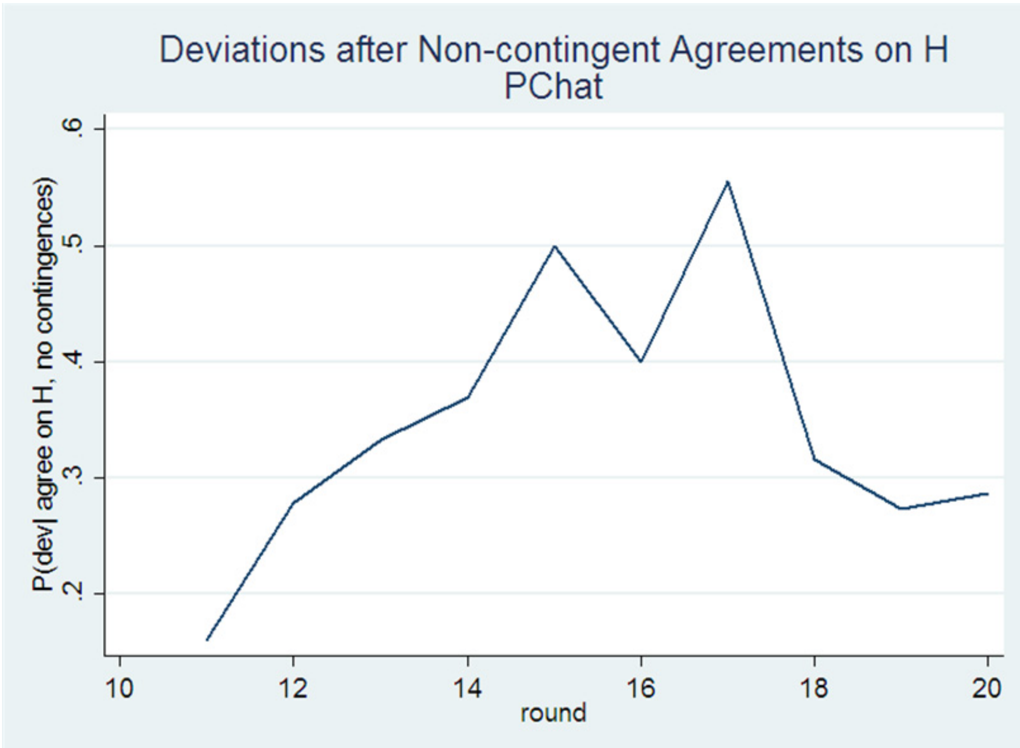

Figure 6

Do Players Learn from Threats?

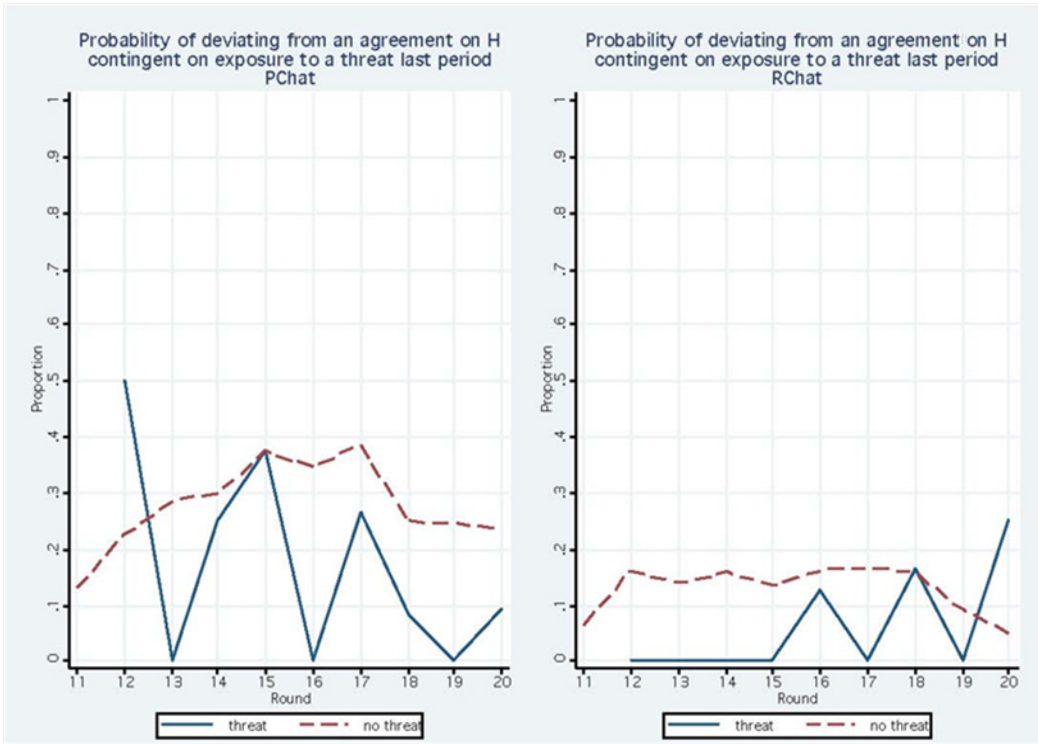

Figure 7

Most Common Period 1 Codes and their Effect

Message Description	Proportion Observed	Average Period 1 Cooperation	
		Not Observed	Observed
Proposed Period 1 Weak Cooperation (Both "Medium")	0.148	0.883	0.576
Proposed Period 1 Strong Cooperation (Both "High")	0.931	0.407	0.858
Proposed Period 2 Cooperation (Both "High")	0.744	0.840	0.831
[illegible]	[illegible]	[illegible]	[illegible]

Effect of Chat on First Period Cooperation

Variable	Model 1	Model 2	Model 3	Model 4
Data Set	All Chat	All Chat	PChat	RChat
# Observations	622	622	320	302
Control for RChat	✓			
Subject Fixed Effects		✓	✓	✓
Period 1 both M	-.203*** (.052)	-.134*** (.053)	-.096 (.067)	-.176** (.071)
Period 1 both H	.213*** (.057)	.237*** (.060)	.244*** (.070)	.247** (.105)
Period 2 both H	.037** (.020)	.025 (.028)	.033 (.065)	.033 (.031)
Disagreement	-.032 (.071)	-.065 (.062)	-.066 (.067)	-.099 (.140)
Agreement	.074** (.031)	.035 (.031)	.057 (.043)	.004 (.047)
Threat of Punishment	.164*** (.024)	.112*** (.034)	.162*** (.039)	.076 (.062)
Appeal to Joint Payoffs	.028 (.020)	.021 (.026)	.000 (.030)	.079* (.012)
Reference to Payoff Table	.010 (.025)	.010 (.025)	.004 (.031)	.029 (.040)
You should be trusted	.077*** (.026)	.067** (.029)	.075* (.039)	.056 (.042)
Lack of trust in other	.018 (.034)	.032 (.043)	.033 (.049)	.043 (.086)
Appeal for trustworthy behav	.016 (.027)	.034 (.031)	.046 (.040)	.038 (.055)
Report of having been cheated	.012 (.026)	-.010 (.031)	.009 (.042)	-.041 (.013)

Table 4

The content of Renegotiation Messages

Message Description	Proportion Observed	Proportion Observed		Group Cooperation (Per. 2 Per. 1 No-Coop)	
		P.1 Cooperation	P.1 Non-cooperation	Not Observed	Observed
Positive Feedback Following Cooperation	0.204	0.283	---	---	---
Apology for Cheating	0.111	---	0.372	0.803	0.842
Rationalizing Cheating	0.103	---	0.339	0.844	0.784
Admonition for Cheating	0.119	---	0.394	0.953	0.697
Proposed Period 2 Cooperation (Both "High")	0.902	0.901	0.906	0.500	0.846
Agreement to Proposal	0.626	0.675	0.511	0.529	0.938
Appeal to Joint Payoffs	0.094	0.017	0.278	0.790	0.856

Table 5

[redacted]

Effect of Punishment on Next Round's Period 1 Cooperation

	Opponent Admonished	Opponent Didn't Admonish
Opponent Used "High"	81%	59%
Opponent Didn't Use "High"	91%	74%

Table 6