Communication, Renegotiation, and the Scope for Collusion

David J. Cooper
Florida State University

Kai-Uwe Kühn
University of Michigan and CEPR

October 15, 2009
Preliminary

Abstract: We use experiments to analyze what type of communication is most effective in achieving persistent collusion in a simple collusion game. Consistent with the existing literature on communication and collusion, even minimal communication leads to a short run increase in collusion. In a limited message-space treatment where subjects cannot communicate contingent strategies, this initial burst of collusion rapidly collapses. When unlimited pre-game communication is allowed via a chat window, an initial decline in collusion is reversed over time. Explicitly threatening punishment of deviations from a collusive agreement is identified via content analysis as the most successful type of message for achieving collusion. This long term return to collusion appears to be tied to increased usage of effective types of communication as the use of explicit threats. Contrary to theory, collusion is even more likely when we allow for renegotiation. What appears critical for the success of collusion is that, cheaters are often admonished in strong terms for their behavior and become less likely to cheat in future rounds after being admonished. Allowing renegotiation therefore increases collusion by allowing for an inexpensive and highly effective form of punishment.

Acknowledgments: We would like to thank the National Science Foundation (SES-0720993) for funding these experiments. Angelo Benedetti, E. Glenn Dutcher, John Jensenius, Cortney Rodet, Micah Sanders, Elena Spatoulos, and Evan Starr provided valuable research assistance on this project. We would like to thank Tim Cason, Guillaume Frechette, Steve Salant, Yossi Spiegel, and seminar participants at Florida State University, IESE, University of Cologne, University of East Anglia, University of Innsbruck, the University of Erlangen-Nuremberg, University of Michigan, the EWEBE meetings in Innsbruck, the IO Society sessions at the ASSA meetings in San Francisco, the Conference on Cartels and Tacit Collusion at Norwich, the IIOC conference in Arlington, the 2009 CEPR Applied IO conference in Mannheim, and the FSU Workshop on Experimental Game Theory for their helpful feedback. The authors are solely responsible for any errors in this manuscript.

"The Finns will respect the Spanish dominance in Spain if ENCE really increase their prices in other countries: If Fincell learn about prices below US $360 also in the future, they will reconsider their policy as to sales in Spain!"[1]

Introduction

Communications containing explicit threats and promises, as in the preceding quote from the European *Woodpulp* case[2], are intuitively understood to be at the heart of illegal cartel activities. Indeed, the per-se prohibition on price fixing under the Sherman Act in the US and Article 81 of the Treaty of Rome in the EU is effectively a prohibition of having such conversations. In contrast, setting prices at collusive levels without explicit communication (i.e. tacit collusion) is widely held to be legal (e.g. Motta 2004, p.189). Antitrust enforcement against collusion therefore focuses on discovering evidence of communication and explicit agreements.[3] Despite this emphasis by anti-trust authorities, incriminating communication commonly occurs within cartels as is well documented in case and field evidence (e.g. Genesove and Mullin, 2001). The sheer frequency of interfirm communication about collusion in the face of large fines suggests that it must be a valuable tool for effectively establishing collusive outcomes. But it is not well understood why this is the case and what exactly needs to be said to make collusion successful.

The theory of repeated games (Abreu, 1988; Abreu, Pearce, Stacchetti, 1990), the standard modeling tool for antitrust policy advice, does not explicitly include communication between firms. However, the theory implies that collusion relies on firms solving a difficult coordination problem since cooperation can only be supported if players know that deviations will be punished by switching from a high payoff to a low payoff continuation equilibrium. The industrial organization literature typically assumes that the necessary coordination on contingent strategies is easily achieved (and hence tacit and explicit collusion are mostly treated as equivalent).[4] However, experimental evidence from closely related coordination games shows that subjects often fail to achieve Pareto optimal outcomes in the absence of an explicit coordination device (Van Huyck, Battalio, and Beil, 1990). Communication dramatically increases the likelihood of an efficient outcome in these simple coordination games (Cooper, De Jong, Forsythe, Ross, 1992; Blume and Ortmann, 2007; Brandts and Cooper, 2007). But it is not obvious that this result extends to collusion games since coordination is far more complex, requiring agreements not just on a single action but on entire contingent plans.

At the same time experimental evidence suggests that communication may be more than just a coordination device. Communication often leads to more cooperative outcomes even when these are not equilibrium outcomes. Examples include social dilemmas (Dawes, MacTavish, and Shaklee 1977), public goods games (Isaac and Walker, 1988), and trust games (Charness and Dufwenberg, 2006). These results suggest that forces such as guilt aversion, lying aversion, and increased group identity allow communication to enhance both trust and trustworthiness, leading to greater cooperation. If this also holds for collusion games, then pecuniary punishments (and by extension contingent strategies) are not required to sustain collusive outcomes and many of the

[1] From document found in the European *Woodpulp* case (see Decision 85/202/EEC (1985)), summarizing a meeting between woodpulp producers in Europe. ENCE is the leading Spanish producer of woodpulp while Fincell is the joint sales organization of the Finnish producers.
[2] Decision 85/202/EEC (1985), published in OJ L85/1.
[3] There is extensive use of "dawn raids" to find such evidence in internal company documents. For example, EC authorities raided EDF in March 2009 to gather information about price fixing in the French electricity markets (Mortished and Pagnamenta, 2009).
[4] See Kühn (2001) for discussion of this issue.

policy implications of traditional collusion theory are questionable. As Whinston (2006) has stressed, an understanding of the role of communication is therefore crucial to properly formulate policies towards cartels and tacitly collusive behavior.

This paper addresses these issues by presenting a series of experiments exploring the relationship between communication and collusion. We investigate the specific mechanism by which communication leads to increased collusion. We also study the persistence of communication's effect on collusion and whether allowing for the possibility of renegotiation changes the impact of communication.

Our subjects play a series of two period collusion games with random rematching between games. The two period collusion game is designed to capture the main strategic features of infinitely repeated collusion games in an environment that is simple to understand and easy to play many times in a laboratory session of reasonable length. Play quickly collapses to the one shot non-cooperative equilibrium when there is no communication.[5] If communication is limited to pre-game statements of intent to collude in Period 1 (and there is *no* possibility of specifying a punishment scheme for deviations from collusion), an initial increase in collusive behavior is followed by a collapse back to the non-cooperative equilibrium. This is in line with results reported by Holt and Davis (1990). When a rich pre-game message space is used – subjects have access to a chat window and can send and receive unlimited messages – there is again an initial burst of collusive behavior followed by gradual deterioration. Unlike the treatment with a limited message space, this decline slows and eventually reverses in the pre-play chat treatment. By the end of the experiment, collusive behavior returns to its initial high levels. If renegotiation is allowed by adding chat between periods of the game, behavior is very collusive and never exhibits a decline. This contrasts strongly with the unambiguous theoretical prediction for the game we implement – renegotiation should eliminate all collusion by making it impossible to credibly commit to punish cheating.

Using detailed analysis of the chat content we can identify the mechanisms by which communication leads to greater collusion. In the treatment with pre-play chat only, the single most effective type of communication for generating collusion is a credible threat that non-collusive play will be punished. Subjects who either send or receive such a message are significantly less likely to cheat on collusive agreements. There are few threats of punishment in early games, but the probability of such threats grows strongly over time. Subjects apparently learn that this approach is effective and therefore adopt it. Statements related to other-regarding preferences are commonly observed in the treatment with pre-game chat only, but they are not as effective as threats of punishment. Receiving (but not sending) a message that appeals to mutual gain to justify collusion leads to weakly less cheating on collusive agreements and sending (but not receiving) a message indicating that the sender should be considered trustworthy is associated with less cheating. In both cases the marginal effect is less than half of the marginal effect of a threat to punish cheating. Appeals for trustworthy behavior, while fairly common, have no effect on the likelihood of cheating. Use of statements related to other-regarding behavior grows over time, but less dramatically than the use of threats of punishment. We consider a large number of other categories of communication, but do not find any that impact the likelihood of cheating.

[5] Payoffs in our game are designed to make it likely that collusion will fail in the absence of communication. In many collusion games some collusion will be observed in the absence of communication. See Dal Bo and Frechette (2008) for an experimental study identifying which features of the payoff table make it likely that tacit collusion will occur.

The preceding observations are broadly supportive of the standard theory of collusion. The theory predicts a specific mechanism underlying collusion, the credible threat of punishment when firms deviate from collusion. Consistent with this mechanism, the most important role played by communication in the absence of renegotiation is to help subjects to coordinate on punishment of cheating. If collusion was observed but credible punishment was rarely mentioned or played, a smaller role in fostering collusion than other types of communication, this would have been a cause for skepticism about the theory. Our result thus complements earlier experimental work on collusion by Dal Bo (2005) which demonstrates that collusive play is more likely when players operate "under the shadow of the future." We show that the "shadow of the future" becomes much more important when *explicit* threats of punishment can be made. As predicted by the theory, the shadow of future retribution then plays a central role in supporting collusion.

Messages relating to trust, trustworthiness, or social welfare only have a limited impact in the treatment with pre-play communication only, and statements of intent to collude quickly lose their effectiveness. These observations may seem surprising given the existing results on communication and cooperation cited above. In fact, consistent with the guilt aversion hypothesis of Charness and Dufwenberg (2006) we find that subjects are less likely to cheat when explicit promises to be trustworthy are made.[6] However this only plays a secondary role in determining the long run frequency of collusion. We conjecture that this lack of long-run impact is driven by two critical features of our experiments. First, subjects play the collusion game many times (with random rematching). In our treatment where only statements of intent are allowed, lying about intent is common even in early rounds. In pairs where both players stated intent to collude, less than half do so in the first game with communication, a proportion that rises with experience. Even individuals who initially tell the truth about their intent to collude will learn that lying is the norm. They should also come to anticipate that others will not believe their statements. The experience of cheating by others should therefore reduce the impact of guilt and lying aversion. It follows that the initial and long-term effects of communication on collusion need not be the same. Second, cooperation is consistent with equilibrium in the collusion game. This gives rise to a highly credible and effective means of calling for cooperation. It is natural that the use of credible threats is the primary mechanism of generating collusion while more subtle statements about trust and trustworthiness play a secondary role.

Analysis of second period chat from the treatment with renegotiation provides a clear explanation for why collusion is more successful than in any other treatment. Consistent with the theory, players try to avoid going through with a punishment following cheating and these attempts have some success. As a result, average monetary punishments after deviations from collusive agreements are the weakest of all communication treatments. However, this appears to be counteracted by a second important effect of allowing to have chat between the two periods of the game: individuals who are cheated can reproach those who cheated them. They seize upon this opportunity with high frequency and great enthusiasm. Receiving such verbal punishments reduces the likelihood of cheating in subsequent games with other subjects, suggesting that subjects find it unpleasant to be chewed out after cheating. The availability of an inexpensive and effective form of punishment in the treatment with renegotiation provides a good explanation for the high and stable levels of collusion achieved in this treatment.[7]

[6] The relevant comparison is to behavior by the same subjects when a promise is not made, not to behavior by subjects who never promise to be trustworthy.

[7] This is similar to the effect of non-pecuniary punishments (disapproval points) in public goods games (Masclet, Noussair, Tucker, and Villeval, 2003). The effect here is more persistent, possibly due to the

Turning to the overall implications of our results, these highlight the importance of making a sharp distinction between explicit and tacit collusion in anti-trust policy. While it is possible to achieve some tacit collusion in the lab, direct communication clearly makes it easier to achieve cooperation. More importantly, our experiments point to the types of communication that should be of particular concern in anti-trust enforcement. Calls for collusion in isolation may not be terribly effective. What truly matters is laying out a punishment for failure to stick to a collusive agreement. The ability to verbally retaliate is also extremely important in our experiment. While it may seem intuitive that infighting among conspirators should be bad for collusion, the desire to avoid such personal conflict may at the same time be a crucial factor in stabilizing collusion. In the conclusion to this paper we discuss whether this result is likely to extend to field settings.

We view our work as a complement to field work studying the transcripts of communication between colluding firms (e.g. Genesove and Mullin 2001). Laboratory experiments have clear advantages and disadvantages relative to such field data. The advantages of laboratory experiments come from avoiding selection effects, control of the environment, and observability of actions. Field data on communication and collusion comes from firms that were sufficiently successful at colluding to warrant prosecution and sufficiently indiscrete (or possibly unlucky) to get caught. In the lab we observe the full population, including firms who try to collude and fail. Field data typically comes from case studies that can be characterized as a single observation of a game (albeit a very rich observation). In the laboratory, we observe many plays of the game including multiple plays by each individual. The controlled environment of the lab allows us to fully observe the incentives faced by subjects, the information they had available, and the full content of their discussions. We can experimentally manipulate what types of communication are available to our subjects. Our experiments cannot match the verisimilitude of field data, but we believe that the richness of the data and the ability to study non-naturally occuring communication structures make them an equally valuable tool for understanding the relationship between communication and collusion.

Our experimental design is motivated by a particular type of cooperation, collusion among firms, but the insights we generate apply more generally. There are many economically relevant situations which share the same underlying supergame structure as collusion models (e.g. team production, provision of local public goods). The main insights from our work are likely to be applicable in these settings as well.

The paper is organized as follows. In section 2 we discuss collusion theory and the theory of communicatgion in games. Section 3 describes the experimental design in detail. In section 4 we present the experimental results. We first analyze the benchmark behavior when there is no communication possible and then look at the short run and long run treatment effects of various communication treatments. We then go into more detail on the content analysis of the pre-play chat treatment and the renegotiation treatment. Section 5 concludes the paper.

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

---

richer set of verbal punishments available. Likewise, Xiao and Houser (2005) find that the possibility of verbal punishment reduces rejection rates in ultimatum games.

long run losses induced by a future switch from collusive to competitive behavior. 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 future 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 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). We use this work to make predictions about the type of communication we should expect if players are to achieve coordination on a collusive outcome.

To better illustrate the theory, we introduce the simple two stage game on which our experiments are based. The first stage of the model consists of a one shot duopoly game which is followed by a coordination game in the second stage. Collusion requires that players use the outcome of the first stage as a coordination device for playing the second stage. This captures the essential structure 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).

The game played in the first period is based on a standard model from oligopoly theory, a symmetric Bertrand duopoly with homogeneous goods. The game is simplified by only allowing 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) shows the payoffs for the stage game:

(1)

Player 1 *payoffs*	L	M	H
L	$\frac{\pi^L}{2}$	π^L	π^L
M	0	$\frac{\pi^M}{2}$	π^M
H	0	0	$\frac{\pi^H}{2}$

The unique Nash equilibrium of the game described in (1) is (L,L). In a typical collusion game we would model the competition between firms as an infinite repetition of the stage game shown 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. To reduce the strategy space while still capturing the essential features of the infinitely repeated game, we instead use the following pay-off matrix for the continuation game, where $\Pi^i = (\delta/(1-\delta))*(\pi^i/2)$ and δ is the discount factor:

(2)

Player 1 *payoffs*	L	M	H
L	Π^L	$\delta[\pi^L + \Pi^L]$	$\delta[\pi^L + \Pi^L]$
M	$\delta\Pi^L$	Π^M	$\delta[\pi^M + \Pi^L]$
H	$\delta\Pi^L$	$\delta\Pi^L$	Π^H

Given the definition of Π^i, the payoff matrix in (2) has three equilibria, in each of which the players choose the same strategy. These equilibria are Pareto ranked with (H,H) being the Pareto

dominant equilibrium. We refer to 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).

The second period game shown in (2) is derived from 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. In the infinite 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) then 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) followed by the most severe punishment equilibrium: When a player is cheated and therefore has a higher price than the other player, he earns zero payoffs in the first period of the continuation game and then $\pi^L/2$ forever. When a player deviates and just undercuts a symmetric equilibrium price at price i, he receives the industry profit π^i in the first period of the continuation game and then $\pi^L/2$ forever.[8]

We assume that the incentive conditions are satisfied so that {(L,L),(L,L)}, {(M,M),(M,M)}, and {(H,H),(H,H)} are subgame perfect equilibrium outcomes of the TSBG if players play (L,L) in the second stage after any deviation in the first. Colluding at either M or H in Period 1 is therefore feasible, allowing us to detect whether communication leads to full collusion or not.

In order for pre-play communication to have a systematic impact on outcomes 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. Two features of communication appear to be critical to have credible communication. First, subjects should reach an agreement on a course of action. If subjects cannot agree on a course of action, it is unclear why one player's proposal should be taken more seriously than the other's. Neither statement should therefore be considered credible. Second, any agreement should be *self-committing* for each player in the sense of Aumann (1990), i.e.each player should have an incentive to do what he agreed to (including in all subgames) as long as he expects the agreement to be believed by the other player.[9] Requiring the agreement to be mutually self-committing is equivalent to requiring agreement on a subgame perfect equilibrium. If players agree on a course of action that is not a subgame perfect equilibrium, both can gain by cheating on the agreement (at least for some history of the game). Given that all subgame perfect equilibria in our collusion game that support Period 1 prices above L involve contingent play in period 2, the theory predicts that communication will facilitate collusion only if messages specifying contingent strategies are available and are used.

[8] The payoff matrix in (2) can therefore be interpreted as a reduced form of the infinite horizon game when attention is restricted to the symmetric optimal punishment equilibria. This is the set of equilibria that is often analyzed in applications of collusion theory in industrial organization.

[9] Another commonly imposed condition for a message to be credible is that it be self-signaling (Aumann, 1990) – the sender only wants the message to be believed if he is telling the truth. Farrell and Rabin (1996) point out that this condition leads to unlikely predictions about the effectiveness of communication in stag-hunt games, a criticism supported by experimental results from Charness (2000). The same issues arise in the TSBG. Calling for play of the collusive equilibrium (Period 1 play of H supported by punishment with L in Period 2) is not self-signaling since a subject who plans on playing L unconditionally in both periods gains by sending this message if it is believed. The argument has little force if players reach an agreement, since playing the collusive equilibrium yields higher payoffs than sticking with L.

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 raises the issue of renegotiation. Suppose in the TSBG the players can communicate before both periods. Once Period 2 arrives both players are left with game (2). Communication is would yield the Pareto optimal equilibrium in this sub-game since an agreement to play this is self-committing. It follows that Period 1 messages about Period 2 actions should be ignored. If messages prior to Period 1 cannot credibly specify a contingent strategy for Period 2, then a message calling for the collusive equilibrium will no longer be credible. No collusion should therefore be possible in the TSBG with the possibility of renegotiation. Put more formally, play of (L,L) in both periods is the only outcome consistent with any of the suggested versions of renegotiation proof equilibrium for finite games.[10]

The prediction that no collusion can be obtained with renegotiation is due to the fact that the continuation game has only Pareto ranked equilibria. The scope for collusion under renegotiation is significantly wider when there are multiple asymmetric continuation equilibria that cannot be Pareto ranked (see Benoit and Krishna 1988 for finitely repeated games and Van Damme 1989, Farrell and Maskin 1989, Abreu, Pearce, Stacchetti 1993 for supergames). For our purposes it is useful to make the predicted effect of renegotiation as stark as possible to get a clean test of the main ideas underlying all models of renegotiation: the ability to renegotiate will eliminate use of Pareto dominated continuation equilibria and therefore limit the outcomes that can be supported as equilibria in the full game.

Our discussion of the theory so far has assumed that the payoff matrices represent a player's utility from outcomes, but there exists a wealth of experimental evidence indicating that this need not hold in laboratory settings.[11] Not only may subjects care about how their payoffs relate to the payoffs of others, putting weight on concerns such as fairness and social welfare, but they also care about the actions that led to the final outcome. Particularly relevant for an experiment on communication, subjects may be averse to telling lies, especially if these lies cause others harm. If lying aversion or guilt aversion are important factors we should see relatively few people cheat on collusive agreements even if there are no threats of punishments or previous experiences of punishment after lying.

Charness and Dufwenberg (2006) report such effects in a trust game with incomplete information. However, there are several reasons why the scope for guilt or lying aversion to affect outcomes in our collusion model may be limited relative to trust games. First, in a trust game the sequential move structure makes it unambiguous for the second mover how much he hurts the first mover by lying. In the TSBG, in contrast, a player cannot know if his failure to follow through on a collusive agreement harms the other player or possibly saves himself from harm due to cheating. If cheating is common (or at least perceived to be common), players should feel little guilt about cheating.[12] Second, cooperation is consistent with equilibrium in our game. Communication

[10] See Bernheim, Peleg, and Whinston's (1987), Bernheim and Ray (1989), and early versions of Farrell and Maskin (1989).

[11] See Cooper and Kagel (2009) for a recent survey.

[12] As modeled by Charness and Dufwenberg, the amount of disutility experienced due to guilt depends on the amount of harm done to the other player and the difference between second order beliefs (beliefs about the other player's beliefs) and the action taken. In the TSBG, if other players are believed likely to cheat, the expected harm from cheating is reduced and the difference between second order beliefs and actions is

about coordinating on such an equilibrium is a possibility that is not available in a trust game but likely to be quite useful in the TSBG. Such communication could therefore crowd out non-contingent promises. Finally, we allow subjects to learn about cheating through repeated play over time. If there is considerable cheating, learning about this fact may lead to a reduction in perceived guilt about cheating over time even for individuals for whom guilt aversion is initially important.

In the context of renegotiation there is a potential for concerns like guilt aversion to reappear in a much more direct form. In particular, renegotiation creates the possibility that an opponent can complain after having been cheated. Even though such verbal punishments have no monetary impact on the recipient, other experimental studies have shown that subjects play more cooperatively when non-pecuniary punishment is possible (Masclet, Noussair, Tucker, and Villeval, 2003; Xiao and Houser, 2005). If sufficiently strong, an aversion to being called out for cheating could lead to higher levels of collusion when renegotiation is allowed rather than lower.

3. Experimental Design

a. General Design

In our view, both collusive strategies and the types of communication that are likely to facilitate collusive behavior require relatively sophisticated reasoning. We do not expect inexperienced subjects to immediately play equilibrium strategies or be able to distinguish between credible and non-credible communication. For this reason our design puts emphasis on allowing the subjects to learn strategic interaction as well as the use of communication in the game. Subjects play twenty rounds of the TSBG in all treatments. A "round" refers to an entire play of the TSBG while a "period" refers to one of the two stage 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 for subjects to fully understand the experimental interface, to be comfortable with the payoff tables, and to 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 that 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 the 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.

b. Use of a Two Period Game

---

also reduced. The same argument has less force in trust games because the second mover knows what the first mover has done and, by extension, has a good idea of whether the first mover has believed him.

By using the TSBG our design differs markedly from earlier work on collusion that has used indefinitely repeated games as proxies for supergames (dal Bo 2005; dal Bo and Frechette 2008; Duffy and Ochs, 2009). Use of a two period game has significant methodological advantages for a study that focuses on communication and learning. Most importantly, it dramatically reduces the complexity of the environment by limiting both the number of actions and the number of periods in the game. The number of potential contingencies that an experimental subject has to consider (and potentially communicate about) is relatively small in the TSBG. If the action space or the number of time periods is increased, the number of potential contingencies increases geometrically. Reducing the complexity of the game helps us fulfill one of our primary goals, understanding what types of communication are most effective for increasing collusion. Analyzing the content of messages is a daunting task even when the number of types of relevant messages that can be sent is small. If we expanded the strategy space and by extension increased the number of relevant message types, this analysis would quickly become intractable.

Simplifying the game to two periods also helps with our goal of observing whether treatment effects are persistent. Subjects have a lot to learn in our design. Not only do they need to learn the main strategic features of the TSBG, but they also need to learn how to send effective messages and how to interpret the messages they receive. To give the learning process a reasonable chance of converging to long run behavior, our experimental design calls for twenty rounds of the game. The first ten rounds are for subjects to learn about the game and interface while the final ten rounds give an opportunity to learn about communication. As will be seen, twenty rounds are far from excessive to observe long run outcomes. For example, we would have missed the most important feature in the PChat data if we had used fewer rounds. We aimed to fit these twenty rounds into a two hour session, including instructions. Based on many years of experience running experiments, we feel two hours is the maximum time a session can last without subject fatigue and boredom becoming significant problems. Open chat slows down play significantly, making it difficult to complete even twenty rounds of play within the allotted time frame. A design in which players played multiple periods (with appropriate discounting) of (1) followed by one period of (2) would have used even more time per round, forcing us to reduce the number of rounds and hence the likelihood of accurately observing long term outcomes. Since the results of a follow-up study (Cooper and Kühn, 2009) indicate that increasing the number of periods prior to the terminal coordination game has virtually no effect on the development of collusion, there is little reason to have fewer rounds in exchange for extra periods in each round.[13]

One simplification that we rejected was having two prices in the stage game rather than three. This simplification would have sacrificed important features of the game. With only two actions it is difficult to distinguish between a failure to coordinate in the continuation game and explicit punishment behavior. With three actions, the second period game can be designed to clearly distinguish punishment. As shown below, the payoffs we use for the TSBG are designed such that M is the likely outcome when the second period game is played in isolation. Given that M is

[13] Alternatively, we could keep the average time of a twenty round session the same while using an indefinitely repeated game by setting a discount rate of $\delta = ½$. But this does nothing to reduce the complexity of the strategy and message space and has the additional problem that half of the rounds (on average) have only a single period. In single period rounds subjects learn nothing about contingent responses to first period behavior, unlike the TSBG where subjects are continually receiving relevant feedback about how first period behavior affects play (and communication in the renegotiation treatment) in the continuation game. Even though the average amount of time spent playing is equivalent, twenty rounds of the indefinitely repeated game with $\delta = ½$ should be expected to yield slower learning and convergence than twenty rounds of the TSBG.

the natural outcome when no coordinating device is available, use of L can unambiguously be interpreted as punishment. Having three prices also allows for an intermediate level of collusion in the first period. Collusion at "Medium" offers some gain over the "Low" outcome with less risk and less need for harsh punishments than collusion at "High".

Although the TSBG captures the main theoretical properties of the underlying supergame, the move from an infinitely repeated game to a finite game potentially has costs. Because the payoff matrices differ for Periods 1 and 2, some subjects may not grasp the relationship between the two periods of the game (or may believe that others will not grasp the relationship). Weakening the link between Periods 1 and 2 should make collusion more difficult to achieve – especially in the absence of communication. Although the data indicate that there is a strong link in all treatments between Period 1 outcomes and Period 2 choices, our approach may nevertheless systematically underestimate the likelihood of collusion. As noted previously, this concern is addressed by a follow-up paper using subjects from Florida State University (Cooper and Kühn, 2009). Among other things, this paper compares play in two and three stage Bertrand games for the treatments with no communication, limited messages prior to Period 1, and open chat prior to Period 1.[14] No differences are observed that would affect the conclusions reported here. If recognizing the link between Periods 1 and 2 was a major problem, we would expect different behavior when the game changed so Periods 1 and 2 were more obviously related.

The simplified environment of the TSBG also restricts the possible methods of building a collusive agreement. 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, are viable possibilities in the supergame but not in the TSBG. However, all of these features would generate additional considerations that would make it harder to cleanly identify the role that communication about contingent behavior plays for achieveing collusive outcomes.

c. The Payoff Matrices in the Two Stage Bertrand Game

In picking specific payoffs for the TSBG we had several goals. First, for reasons described previously we wanted to make (M,M) focal for the Period 2 game played in isolation (i.e. when play is *not* contingent on Period 1 outcomes). We hoped to achieve this by making the equilibrium (M,M) risk dominant (see Harsanyi and Selten 1988). Second, we did not want reversion in Period 2 from the (H,H) equilibrium to the (M,M) equilibrium to be a sufficiently strong punishment of Period 1 cheating to support collusion at H in Period 1. In other words, we wanted collusion to be attractive but only achievable if subjects are willing to use a strong punishment. This feature helps us to distinguish punishment from coordination failure. Finally, we wanted to make the payoff from colluding at H in both periods sufficiently large relative to the payoff from deviating to M in Period 1 that subjects would have a clear incentive to follow a collusive agreement.

To achieve these two 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.

[14] The renegotiation treatment takes too long with the three period game to complete twenty games in a two hour session.

(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) in both periods is 25% greater than the payoff from defection to M in Period 1 and reverting to the (L,L) equilibrium in Period 2 (180 vs. 144). Play of (M,M) in Period 1 can also be supported as a collusive equilibrium. This requires less in the way of punishment – reversion from (H,H) to (M,M) in Period 2 is sufficient – but obviously yields a lower payoff than colluding at H. Parameters have been chosen such that (M,M) is risk dominant in Period 2 relative to (L,L) and (H,H).

d. The Communication Treatments

Our experimental design compares play in four communication treatments: No Communications, Period 1 Limited Communication, Pre-play Chat, and Renegotiation. In the following subsection we briefly describe each of these treatments.

No Communication (N Treatment):

The N treatment serves as a benchmark for treatments with communication. The rules for Rounds 11 – 20 are identical to those in Rounds 1 – 10, with no communication between players allowed. The N 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. The N treatment therefore acts as a control for restart effects.

Period 1 Limited Communication (P1 Treatment):

The P1 treatment gives subjects the opportunity to send a message prior to the beginning of Period 1. The message space is limited to suggesting actions for Period 1. Specifically, subjects are given the prompt, "I think we should choose the following in Period 1." They are asked to choose between "Low", "Medium", "High", or "No Response" for both "My Choice" and "Your Choice." Messages are chosen simultaneously and each player is shown both parts of both players' messages at the same time as choices are made for Period 1. The feedback at the end of Period 1 reiterates the messages as well as reporting the outcome for Period 1. The critical feature of this treatment is what is not allowed – subjects cannot send any messages about their intents for Period 2. The purpose of this treatment is to establish that simply calling for collusive

behavior in Period 1 is not sufficient to generate Period 1 collusion. The P1 treatment gives a baseline that the effects of rich communication can be compared with.

Pre-Play Chat (PChat Treatment):

At the heart of our experimental design are the two chat treatments. Starting in Round 11, 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 minimal limits on what they can say (subjects are asked to avoid offensive language and identifying themselves). Subjects are given no guidance on how the chat should be used or what they might say, although it is fairly obvious that it is meant for discussing the game.

Renegotiation (RChat 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. . This allows 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 uses the same communications technology as PChat.

Initial Hypotheses: Our initial hypotheses are driven by the discussion of communication and collusion contained in Section 2. While it is possible to achieve tacit collusion in an experimental collusion game, the TSBG is designed to make this unlikely.[15] The limited message space allowed in the P1 treatment makes it impossible to reach a mutually self-committing agreement to collude in Period 1. We therefore did not expect long term collusion in this treatment. Given that it might take some time for subjects to figure out that calls for collusion are not credible, we expected some collusion initially upon the introduction of messages. This would be consistent with the earlier results of Holt and Davis (1990) who find that communication about actions (but not about contingent strategies) leads to a transient increase in collusion in a posted-offer triopoly. While guilt aversion could lead to increased collusion in the P1 treatment, as compared to the N treatment, the results of Charness and Dufwenberg (2008) indicate that the effect is likely to be small in an environment with such a limited message space. For the PChat treatment we expected to see a persistent increase in collusion beyond what is observed in the N and P1 treatments since it is possible to reach agreements to collude that are self-committing. The rich communication available in this treatment also provides more scope for behavioral factors such as guilt aversion to have an effect. The theory of renegotiation (Bernheim, Peleg, and Whinston, 1987; Farrell and Maskin, 1989) predicts that allowing chat between Periods 1 and 2 should

[15] The TSBG is designed so collusion at H is not risk-dominant. The results of Dal Bo and Frechette (2008) indicate that making collusion at H risk-dominant is necessary but not sufficient to guarantee successful collusion. We therefore did not expect collusion at H. *Ex ante* we had no strong expectations about whether play would settle at M or L.

render agreements to collude at H in Period 1 supported by punishment with L in Period 2 non-credible. We therefore expected to see less collusion in the RChat treatment than in the PChat treatment. This prediction is mitigated by the possibility that verbal punishments could serve as an inexpensive means of supporting collusive agreements.

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.[16]

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 to 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.

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

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

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 N treatment with a short announcement that none of the rules would change.

In the sessions with the P1 treatment, 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.[17] 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.

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

[17] 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.

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. With one exception, the content of the chat contains no evidence that subjects knew when they had played an opponent previously.[18]

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.[19] 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.[20]

4. Results

Collusion in the TSBG is defined in terms of Period 1 play. When we discuss collusion in the results section, we always refer to Period 1 choices. Our analysis of Period 2 choices focuses on how Period 2 choices depend on Period 1 outcomes. This is what is of central interest about Period 2 behavior for distinguishing between various theories of collusion, communication, and renegotiation.

A) Behavior in Periods 1 -10: Our experimental design includes in all treatments an initial phase of ten rounds without communication. Subjects in all treatments saw initially modest levels of Period 1 cooperation collapse over time. In Round 1, 44% of Period 1 choices were of the least cooperative action, "Low". Use of the most cooperative action, "High", was rare even in the first round, accounting for only 12% of observations. By Round 10, play had converged strongly to the Period 1 stage-game Nash equilibrium with 86% of subjects choosing "Low" for Period 1 and only 4% choosing "High". Differences between treatments in Period 1 Cooperation for the first ten rounds are small and not statistically significant.[21] This is as expected since treatments are identical for Rounds 1 – 10. It is unsurprising that Period 1 cooperation collapses given the incentives faced by subjects. The sum of a subject's payoff for Periods 1 and 2 following Period 1 choices of "High", "Medium", and "Low" average 34, 82, and 98 ECUs respectively for Round 1, with similar incentives (40, 69, and 81 ECUs) over Rounds 1 – 10.

[18] One subject in the RChat treatment passed on the identity of a subject who had cheated him. This had minimal impact – the subjects who were warned only met the offending party in two observations. Not surprisingly, none of our conclusions are affected if these two observations are dropped.
[19] One session of the N treatment only has nineteen rounds of data due to a software problem that caused Round 20 data to not be saved.
[20] We had no specific reason to expect contamination, but were unusually careful because CWRU has a relatively small undergraduate population (about 3500 students at the time).
[21] This statement is based on ordered probit regressions with the Period 1 choice as the dependent variable. Using the N treatment as the base, none of the three treatment dummies are statistically signficant at the 10% level, a result that does not change if one of the other treatments is used as the base.

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 Period 2 choices. Period 2 choices are dependent on Period 1 choices: aggregating across Rounds 1 – 10, if the other player's Period 1 choice is "High" or "Medium", then "High" is chosen in Period 2 for 40% of observation versus 16% choosing "Low". However, if the other player chose "Low" in Period 1, the distribution of Period 2 choices shifts to 26% choosing "High" and 28% choosing "Low". This effect of Period 1 play on Period 2 choices strengthens over time, but is never sufficiently large to prevent "Low" from being the profit maximizing choice for Period 1. The degree to which Period 2 behavior is contingent on Period 1 outcomes suggests that the type (although not the magnitude) of contingent behavior needed for tacit collusion to emerge is present in the data.

The data also contains weak evidence of learning in response to punishment. Consider subjects who did not collude and faced an opponent who tried to collude and then punished (i.e. subjects who chose "Low" in Period 1 and faced an opponent who chose "High" in Period 1 and "Low" in Period 2.). In the following round, only 55% of these subjects choose "Low" as opposed to 79% for subjects who saw the same Period 1 outcome but were not punished in Period 2. This difference is not statistically significant, in part because the relevant sample is tiny (96 observations overall, only 11 with "Low" chosen in Period 2). This illustrates a general problem for the emergence of tacit collusion in our experiment: subjects learn in a way that could lead to collusion, but generally do not receive strong enough feedback to learn to collude tacitly.

Regularity 1: Play in Rounds 1 – 10 collapses to non-cooperation. The type of contingent play that can lead to tacit collusion is present in this initial phase, but is too weak to overcome undercutting incentives.

B) Prices in Rounds 11 – 20: Figure 1 shows that introducing communication dramatically changes behavior. 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 a minimal restart effect with a slight increase for Round 11 in play of Medium relative to Low. The evolution of play then continues the pattern from Rounds 1 – 10, with steady movement toward 100% play of Low in Period 1. In contrast, all three communication treatments show a large increase in Period 1 cooperation as choices of High and Medium increase sharply in Round 11 when communication becomes available with a corresponding drop in use of Low. After this initial increase the evolution of play through Rounds 11 – 20 differs sharply across the three communications treatments.

[Figure 1 about here]

The P1 treatment shows the least initial increase in cooperation of the communication treatments, with Medium being the modal choice in Round 11. Over the following rounds there is a sharp decline in the use of High, which almost dies out by round 20, as well as a steady decline in the use of Medium. Play has not converged by Round 20, but is obviously headed toward the non-cooperative outcome. It is unlikely that there is any substantial difference between long run behavior in the P1 and N treatments.

For both chat treatments in Round 11, the most popular Period 1 choice is overwhelmingly High. Use of Low is almost non-existent for both chat treatments in Round 11. The subsequent dynamics differ dramatically across the two chat treatments. In the PChat treatment there is a distinct decline in cooperation at High until Round 16. This is similar but not identical to the decline observed in the P1 treatment, consisting of movement from High to Medium rather than from High and Medium to Low. Unlike the P1 treatment this decline is completely reversed in Rounds 17 – 20. Round 20 choices are largely indistinguishable from those made for Round 11 with 72% of the subjects choosing High and 56% of the pairs successfully colluding at High (i.e. both players choose High in Period 1). There is a clear trend towards collusion in the long run.

In the RChat treatment, no decline in cooperation is observed. The distribution of actions is essentially constant over Rounds 11 – 20. The use of Low is negligible throughout. The small shifts upward in Rounds 17 and 19 are artifacts driven by RChat sessions ending in different rounds. Breaking the data down by sessions, little change is observed over 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.

Appendix A reports the results of two ordered probit models that provide statistical backing for the proceeding observations about Period 1 prices in Rounds 11 – 20. The first model finds that the described differences *between* treatments are statistically significant. Specifically, immediately after the introduction of communication Period 1 prices are significantly higher in the P1 treatment than in the N treatment, and higher in the PChat treatment than in the P1 treatment. No significant difference is observed between the PChat and RChat treatments in Rounds 11 and 12. The same ordering across treatments holds throughout Rounds 11 – 20, with all differences between treatments statistically significant in Rounds 13 – 20. The second model looks at the changes *within* treatments over time. Most importantly, it confirms that Period 1 prices decline and then rebound significantly in the PChat treatment. In other words, the dip and recovery in the PChat treatment are unlikely to have occurred by chance. Both models correct for individual effects by including a variable measuring cooperative behavior in Periods 1 – 10 as well as correcting standard errors for clustering at the subject level.

Regularity 2: Play in all three communication treatments shows a sharp increase in collusion for Round 11. In the P1 treatment, cooperation collapses over subsequent rounds, returning to the non-cooperative outcome. In the PChat treatment, cooperation initially collapses, but then reverses itself and returns to its Round 11 levels. Both the initial decline and the recovery are statistically significant. Cooperation is high and weakly increasing in the Rchat treatment. The theory correctly predicts the lack of a long-term improvement in cooperation for the P1 treatment, but fails to predict the high degree of collusion in the Rchat treatment.

C. Content Analysis

To systematically study how the content of chat affects play in Period 1 and Period 2, we quantified content by coding all of the dialogues. This process began by developing a coding scheme using a test sample of messages. The goal was to have a code for any type of message that might be relevant for the play of the game. We did not limit codes to categories that we thought were likely to be important in generating collusion, choosing instead to be as inclusive as possible and then let the data decide which sorts of messages are important and which are not. For more details on what categories were coded as well as how the coding scheme was developed and implemented, see Appendix B.

Two research assistants, one at Michigan and one at Florida State independently coded all messages. 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 that were to be tested. 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 codes for a message – a coder could check as many or few categories as he deemed appropriate. Cross-coder correlations for major categories averaged .66, about the same as earlier studies using the same methods.

Table 2 summarizes the frequency of the most common categories, defined as any category that was coded in at least 10% of the dialogues for either the PChat or RChat treatments.[22] Implicit threats are included as a category of particular interest, even though this category does not reach the 10% threshold in either chat treatment. For codes that relate to chat prior to Period 1, shown in the top panel of Table 2, frequencies are reported separately for the PChat and RChat treatments. Codes that relate to chat prior to Period 2 (and therefore must come from the RChat treatment) are shown in the bottom panel of Table 2. They are broken down by whether players successfully colluded in Period 1 (both players choose High) or one player cheated (deviated from an agreement to choose High) while the other did not.[23] The final column of each panel shows κ, a common measure of inter-coder agreement. The κ's generally show substantial agreement between the coders, especially since the number of possible codes was large and no attempt was made to force agreement between coders.

[Table 2 about here]

Table 2 reveals a number of features of communication in our setting. Subjects almost always engaged in communication relevant to the game. For example, Period 1 prices were proposed in 618 of 622 dialogues.[26] The content of the dialogues differed substantially between the PChat and RChat treatment. Some of these differences between the treatments are to be expected. For example, subjects are much less likely to propose a Period 2 action prior to Period 1 in the RChat

[22] To be precise, this is the average percentage across the two coders.
[23] There are a small number of observations where both players cheated that are not included here.
[26] Only one of the remaining four dialogues had no substantive discussion prior to Period 1.

treatment. This makes sense since subjects will have another opportunity to discuss Period 2 prices prior to Period 2. Other differences are more striking:. Subjects in the PChat treatment were far more likely to make a threat to punish cheating with play of Low in Period 2 and all the categories that can be interpreted as relating to other-regarding preferences, such as discussions of trust and joint payoff maximization, are more frequent in the PChat treatment. Subjects not only behave differently in the two chat treatments, they also communicate differently. In acknowledgement of this pattern, we do *not* pool data from the two chat treatments in the content analysis

The first step in the content analysis is to document how often subjects reach agreements to collude, how often they cheat on these agreements, and what the (average) incentives to cheat is. Then we examine two classes of period 1 communication that are both common and may plausibly affect the likelihood of successful collusion: threats of punishments and statements relating to other-regarding preferences. We estimate the impact of these types of period 1 communication on the likelihood of cheating in both treatments. We then briefly examine to what extent different communication strategies are learnt and show that cheating incentives significantly decrease when some types of communication are received or sent. This is especially the case for threats. We conclude the content analysis with a study of second period chat in the PChat treatment, which proves critical to explaining the outcomes in this treatment.

a. Agreements and Cheating

In all three communication treatments we can identify agreements between the players on what price to set in Period 1.[27] For the P1 treatment, we define the players as having come to an agreement if their messages prior to Period 1 suggest the same Period 1 prices. In the PChat and RChat treatments, the players are defined as having an agreement if a proposal for Period 1 prices is made by one player and then accepted by the other. The agreed price is given by the *last* proposal that was agreed upon.

Figure 2 plots the percentage of agreements on Low, Medium, and High over time for the three communication treatments. To make it easier to see the probability of some agreement being reached, the probabilities of each type of agreement are stacked. For example, in Round 11 of the P1 treatment, the probabilities of agreeing on Low, Medium, and High respectively are 0%, 9%, and 53%. The probability of some agreement being reached is 62%, the sum of these three probabilities.

[Insert Figure 2]

Players almost always reach an agreement on Period 1 prices in the two chat treatments, but not in the P1 treatment. In the chat treatments the players can iterate back and forth if they do not immediately reach an agreement, but the limited communication in the P1 treatment does not

[27] In theory the players could agree to choose different prices for Period 1 (e.g. I'll choose High and you should choose Medium), but in practice they do not. The same Period 1 price is specified for both players in all of the agreements we identify in the data.

allow this possibility. This raises a potential problem with our interpretation of the data. We hypothesized that there would be less Period 1 collusion in the P1 treatment than in the two chat treatments because it is not possible to propose contingent strategies for Period 2, but another possibility is that there is less Period 1 collusion because players fail to reach agreements on Period 1 prices. Fortunately, as will argue below, the data does not support this explanation.

Subject to reaching an agreement, players in all three treatments with communication overwhelmingly agree on choosing High in Period 1. This may seem strange in the P1 treatment since players rarely choose High in Rounds 16 – 20. As suggested by the theory of cheap talk, calling for play of High in Period 1 is not credible in the P1 treatment, but it remains rational to announce High as long as there is a sufficiently high probability that some opponent takes this at face value. Indeed, for Rounds 16 – 20 of the P1 treatment, subjects who lie by announcing High but choosing Low earn on average about 10 ECUs more over Periods 1 and 2 than subjects who tell the truth by announcing and choosing Low (83.3 ECUs vs. 73.5 ECUs).

Figure 3 shows the proportion, by treatment, of subjects cheating (choosing Low or Medium) on agreements to choose High in Period 1. The probability of cheating in the P1 treatment is high even in early rounds and rises steadily: In Round 11, 48% of the players who reach some collusive agreement (Medium or High) cheat on it (i.e. choose a lower than agreed price), with the likelihood of cheating being slightly higher (53%) if the agreed on price is High. The likelihood of cheating on a collusive agreement rises steadily over time, with 86% of players who reach some collusive agreement in Rounds 16 – 20 cheating. The high frequency of cheating in the P1 treatment suggests that the cause of declining collusion is *not* a failure to reach agreements. Even if all players reached a collusive agreement, it is unlikely that collusion could survive such pervasive cheating. Consistent with this argument, Period 1 prices by players in the P1 treatment who reach a collusive agreement trend down just as strongly over time as Period 1 prices for players who fail to agree.

[Figure 3 here]

For the RChat treatment, cheating is low and steady throughout. More interesting is the PChat treatment. Cheating initially is only slightly higher than in the RChat treatment, but rises steadily to a peak in Round 17. Cheating then declines back to almost its intial levels. This u-turn in cheating mirrors the turnaround in Period 1 prices for the PChat treatment. An explanation for why Period 1 prices rebound in the PChat treatment therefore must also explain the rise and fall of cheating in this treatment.

Having identified when players have reached agreement as well as when they cheat on these agreements, we can start answering a critical question: Why do subject cheat on agreements to collude in Period 1? There is always a potential incentive to deviate from an agreement on mutual play of High in Period 1 due to the gain in short term profits. Additionally, the gain from cheating increases when subjects perceive a high probability of cheating on the agreement by the

player they are facing.[28] This gain from cheating is reduced to the extent that players who comply with a collusive agreement react punish deviations by their opponent, i.e. the extent to which opponenets are expected to play contingently. Beyond the financial incentives to stick to a collusive agreement, there may be psychological benefits as well since subjects may dislike inequitable payoffs or breaking their word. We cannot directly quantify the psychological costs of cheating, but we can examine the financial incentives to cheat. To the extent that financial incentives do a poor job of explaining when cheating occurs, this suggests that psychological factors may be at play. The content analysis will help us identify what these other factors might be.

The high degree of expected cheating by others provides an obvious explanation for why collusion vanishes in the P1 treatment. Looking at Figure 3 again, the initial level of cheating on collusive agreements to play High in Period 1 is 48% in the P1 treatment. If the other player cheats, a player who does not cheat loses 76 ECUs on average versus cheating. This type of frequent, large loss makes it very difficult to not cheat even if your initial inclination was to stick to collusive agreements and punish cheating in a manner consistent with equilibrium. This force is much weaker in the two chat treatments since the rates of cheating are far lower in the early rounds.

Figure 4 examines how much players are punished for unilateral deviation from a collusive agreement (i.e. cheating by one player but not the other). The data for this figure is taken from subjects who reach an agreement to play High in Period 1 and do not cheat on the agreement. The graph plots the difference between the probability of playing High in Period 2 when the other player does not cheat on the agreement and the same probability when the other player cheats. Higher values of this difference indicate larger responses to unilateral cheating. This variable can be interpreted as measuring the degree of contingency in second period play. The data is grouped into two round chunks – combining rounds reduces the pure noise in the graph by increasing the amount of data in each cell. The markers for cells that have five or fewer observations are displayed as hollow rather than filled in.[29] Two cells in the P1 treatment have no observations with both players sticking to an agreement to choose High in Period 1 and therefore no marker is displayed.

[Figure 4 here]

In all treatments unilateral cheating is punished, with the players who are cheated choosing High less frequently in Period 2 than when they are not cheated. Surprisingly, punishment of unilateral

[28] The subgame perfect equilibrium that supports collusion at High in Period 1 does not limit what equilibrium is played for Period 2 if both players deviate in Period 1. Suppose that the (Medium, Medium) equilbrium is played. This is the risk dominant equilibrium for the Period 2 game and also the modal outcome following mutual cheating. Suppose a player's opponent deviates from the subgame perfect equilibrium by playing Medium in Period 1. If the player sticks to the equilibrium strategy, his payoff over the two periods is 6 ECUs. If he also deviates by choosing Medium, his payoff increases to 135 ECUs.

[29] In the P1 treatment, the hollow marker is a cell where there are fewer than five observations where both players stick to an agreement to choose High in Period 1. The hollow marker in the RChat treatment is a cell where there are fewer than five observations with unilateral cheating.

cheating is strongest in the P1 treatment. The type of contingent behavior needed to support collusion is clearly present even though collusion rapidly collapses because of the high expected rate of cheating. The amount of punishment is initially about the same for the two chat treatments. However, in the PChat treatment the degree of contingency increases strongly over time, while it decreases in the RChat treatment. The level of punishment is very low in the long run for the RChat treatment. This is consistent with the logic of renegotiation theory which predicts that allowing for renegotiation limits the threat of punishment in Period 2.

Putting together the incentives to deviate in period one and the punishments of period 2, Figure 5 shows the (average) net gains from cheating on an agreement to play High in Period 1. This graph plots the difference between total payoffs over both periods when a player cheats on an agreement to play High in Period 1 and total payoffs when a player does not cheat. Higher values indicate larger gains from cheating. The dashed line shows a natural benchmark for gains from cheating: the increase in total payoff over both periods from cheating versus not cheating assuming the other player follows the agreement to play High in Period 1 and plays High in Period 2 regardless of the Period 1 outcome (i.e. doesn't punish in Period 2). If a player does not expect to be cheated, this is the most he can hope to gain by cheating.

[Figure 5]

In the P1 treatment, cheating is always fairly profitable and sometimes even more profitable than in the benchmark because of the high rate of cheating. In the PChat treatment, it is initially more profitable to cheat on an agreement than to follow it, but over time the incentive to cheat becomes much smaller (without disappearing completely). The opposite trend emerges in the RChat treatment. In early rounds, there is little difference in the profits of those who cheat and those who do not. Over Rounds 14 – 20 this changes as cheating becomes clearly more profitable. While there is little cheating on agreements to choose High in Period 1 for the RChat treatment, there is not enough contingent behavior in Period 2 to make this compliance consistent with financial incentives.[30]

Regularity 3: Behavior in Period 2 is strongly contingent on Period 1 outcomes for all three treatments. Only in late rounds of the PChat treatment and early rounds of the RChat treatment is this response sufficiently strong and cheating by the other player sufficiently unlikely that cheating is not clearly profitable.

The analysis of cheating and punishment leaves us with three important points. First, the strong incentives in the P1 treatment provide a clear explanation for why the initial increase in collusion when communication is introduced quickly dissipates. There are two observations that are somewhat puzzling. First, neither chat treatment are the average gains from cheating negative at any point, yet both treatments show high levels of collusion. Second, the incentives to cheat are fairly high in the long run for the RChat treatment, particularly in comparison to the PChat

[30] Econometric analysis, available from the authors upon request, confirms that these patterns (cheating is strictly profitable throughout in the P1 treatment, only in early rounds for the PChat treatment, and only for late rounds in the RChat treatment) are statistically significant.

treatment, yet this is the treatment with the highest and most stable levels of collusion. Taken together, these points suggest that something in the messages must be counteracting the poor incentives to not cheat on collusive agreements. We therefore turn to the effects of two major classes of messages – messages that threaten to punish deviations from a collusive agreement and messages that somehow draw on other-regarding preferences to promote collusion.

b. Credible Messages and Punishment

According to the theory of cheap talk, an agreement to play High in Period 1 is only credible if accompanied by a suggestion for contingent Period 2 play that makes choice of High incentive compatible in Period 1. Specifically, there needs to be a proposal that collusion in Period 1 (i.e. mutual choice of High) leads to mutual play of High in Period 2 while any other outcome in Period 1 will be punished by play of Low in Period 2. In this subsection we examine the incidence and effect of Period 1 messages that include threats to punish cheating. We also compare the behavior when such threats explicitly spell out the punishment scheme as opposed to cases where the threat is non-specific.

Our coding scheme distinguishes between two different types of proposed punishment schemes, explicit and implicit. Explicit punishment refers to cases where subjects specifically suggest that failure to collude will lead to use of Low rather than High in Period 2. Often times this proposal includes an explanation of why the threat of punishment makes collusion incentive compatible. The following dialogue from a game in the PChat treatment is a particularly nice example of this type. The leading numbers indicate which player is talking. This is a verbatim transcript, dodgy grammar and spelling included, except we have replaced the abstract labels subjects saw for prices with the more descriptive terminology used elsewhere in this paper:

1 ^*%&(*^)&(*^(%&
2 im starting to hate people lol
1 ya i know the feeling
2 do you?
1 there are some real jerks out there
2 bcs so did the last person who skrewed me lol
1 haha
2 so whats the plan?
1 [Period 1 High] then [Period 2 High]?
2 yea. if you but [Period 1 Medium] tho i'm putting [Period 2 Low] for the nxt one.. lol.
1 why would you do that?
2 [Period 1 High] and [Period 2 High] ... yes...
1 why would you put anything other than [Period 2 High] for the last one?
this game isn't dependant on how bad other poeple do
you are giving up money just to spite someone you will never know
2 if you skrew me on the 1st one im skrewing you bak no matter what thats why lol
1 that doesn't make any sense
2 lol so you are planning on putting be

*[Period 1 Medium]
1 of course
2 alright
1 ... lol....
2 dont be stupid
bcs
i will. put . [Period 2 Low]
1 since i know you are going ot screw me on the second one
i'll put [Period 1 High] then [Period 2 High]
no point in doing anything else
might as well get 180
2 i'm choosing
1 right .

Subject 1 does not initially understand why Subject 2 would not choose High in Period 2 after cheating, so Subject 2 makes it clear that he will punish cheating. Subject 1 eventually agrees to choose High in Period 1 so as to avoid punishment. Indeed, these subjects successfully colluded.

Implicit punishment refers to cases where a subject threatened to punish non-collusion but did not specify how this would work.[31] The following dialogue gives a typical example of an implicit threat to punish:

1 [Period 1 High]/[Period 2 High]?
2 lets call it a truce: [mutual play of Period 1 High], if you play me all bets are off, and vice versa
2 [mutual play of Period 1 High], [mutual play of Period 2 High]
1 sounds good
1 submitting [Period 1 High] then

There is a clear threat to punish in this dialogue, but no detail about what that punishment might be. Such a message should not be credible since it is unclear that a player is worse off by cheating and then getting punished. For this particular dialogue, the implicit threat worked since the two subjects colluded successfully.

[Figure 6]

The upper left panel of Figure 6 shows the frequency of threats in the PChat and RChat treatments. Data has been grouped into two round chunks to reduce the noise for all panels of Figure 6. Solid lines show the percentage of dialogues (averaged across the two coders) where an explicit threat was observed, and dashed lines show the percentage of dialogues where either type of threat was observed. The difference between the solid and dashed lines for each treatment shows the percentage of dialogues where an implicit threat was made but an explicit threat was

[31] The coding scheme also includes a category for threats to punish with Medium rather than Low, but there were no observations for which this was coded.

not.[32] The use of threats rises over time in both treatments. The frequency of threats is always higher in the PChat treatment than in the RChat treatment, and most threats are explicit in the PChat treatment but not in the RChat treatment.

The remaining panels of Figure 6 examine the effect of threats on the incidence of cheating. Data is included from all observations, with subjects automatically classified as not cheating for the rare cases where no agreement is reached. Hollow markers indicate cells with less than five observations. Only implicit threats are considered for the RChat treatment, since too few explicit threats are made in this treatment to draw any conclusion about their effect.

The top right panel plots the percentage of subjects in the PChat treatment who cheat on agreements as a function of whether they received or sent an *explicit* threat in the dialogue preceding Period 1.[33] Two points are worth noting. Cheating on an agreement is always less frequent when an explicit threat is received or sent. The effect on cheating of receiving an explicit threat grows over time. Note that there is also a late round reduction of cheating even when explicit threats are neither received nor sent. The use of explicit threats can therefore not be the exclusive explanation for the turnaround in aggregate collusive behavior in the PChat treatment. We return to this point below.

The bottom left and right panels show the percentage of subjects in the PChat and RChat treatments, respectively, cheating in Period 1 as a function of whether an *implicit* threat was received or sent prior to Period 1. The graph for the RChat treatment is difficult to read for Round 15 – 20 because two of the lines are on top of each other. For these rounds, no cheating is observed either by subjects who send or receive an implicit threat. The effect of implicit threats differs across the two chat treatments. Focusing on Rounds 15 – 20, the rounds where implicit threats achieve reasonably frequent use, receiving an implicit threat has no obvious impact on the probability of cheating in the PChat treatment. This is consistent with the theoretical prediction that only threats that include a specific punishment should be effective. Sending an implicit threat is associated with a clearly lower frequency of cheating in the PChat treatment. This may reflect sloppiness on the part of senders – subjects sending an implicit threat may well have punishment with Low in mind, but simply fail to communicate this via their messages. Receiving an implicit threat reduces cheating in the RChat treatment, as does sending an implicit threat. Our specific discussion of communication in the RChat treatment, contained in subsection 4Cd, examines why implicit threats are more effective in the RChat treatment.

c. Other-regarding Preferences

The theories of collusion and cheap talk make use of threats a natural focus for our analysis of chat content. However, the literature on communication and cooperation provides alternative

[32] Both types of threats were coded for about 4% of the dialogues in the PChat treatment and 2% of the dialogues in the RChat treatment. Coding of both types of threats in a dialogue largely comes from cases where a subject clarified an implicit threat by subsequently making an explicit threat.
[33] A message is categorized as including a threat if this was coded by *either* coder. This rule is followed in all figures where the presence or absence of a code must be forced to be a binary variable.

explanations for why communication may be an effective tool for generating cooperation. In particular 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. More generally, communication might serve to reinforce norms of other-regarding behavior. Given that the late round improvement in the PChat treatment cannot be entirely explained by the adoption of explicit threats and that behavior in RChat cannot be explained by the use of explicit threats at all, communications related to other-regarding preferences and behavior are natural candidates for an alternative path to collusion in Period 1.

Three specific code categories obviously invoke other-regarding preferences: explaining proposed collusion by appeals to joint payoffs, indicating that you should be trusted, and appeals for trustworthy behavior by the other player.[34] The following example is typical of appeals to joint payoffs.

1 if we both choose [Period 1 High], we can get the most money
2 [mutual play of Period 1 High and then], [mutual play of Period 2 High]. I'm up for it
1 yup
1 i'm choosing
2 me 2

The other-regarding aspect of the dialogue is subtle, lying in the stress on *mutual* benefit – one player is explaining to the other that if *both* players cooperate then *both* players will make more money. In this particular case the two subjects succeeded in colluding for Period 1. Previous work on coordination games (Brandts and Cooper, 2007) found that comments stressing mutual gains were helpful in overcoming coordination failure. Given the relationship between coordination games and supergames, there is some reason to believe this type of comment will help here as well. In general, work on other-regarding preferences has found evidence of preferences for social efficiency (Charness and Rabin, 2002; Engelmann and Stroebel, 2004).

More obviously appealing to other regarding preferences are comments indicating that you should be trusted and appeals for trustworthy behavior. The following is a good example of a dialogue coded under both of these categories.

1 Hi
2 hi
2 [Period 1 High] and [Period 2 High] work?
1 Let's put [Period 1 High] and [Period 2 High]
1 Yeah
2 ok
2 ppl have been so tricky
2 i mean honestly
1 YEah

[34] These are not the only coding categories related to other-regarding preferences. We focus on categories that are frequently coded and are intended to evoke greater cooperation (as opposed to expressing distrust).

2 so PROMISE?
1 I swear
2 haha not that we could kno anyway
2 but it's like ppl don't have consciences...
2 k clicking now

Following a proposal of cooperation in Periods 1 and 2, one of the players indicates concern about being tricked (without specifically accusing the other player) and then requests a promise. The other player responds by making the requested promise. The first player clearly does not trust this promise – a well-founded fear as it turned out. The player requesting a promise chose High in Period 1 but the player who made the promise was lying and chose Medium. As shown here, statements of trustworthiness and appeals for trust often take the form of promises. As a shorthand, we refer to statements that indicate a subject should be considered trustworthy as "promises of trustworthy behavior" even when an explicit promise is not included.

[Figure 7 here]

The upper left panel of Figure 7 shows the percentage of dialogues (averaged across coders) in the PChat treatment where these three categories were observed. Again, data from all observations is included and data is grouped into two round blocks to reduce noise. Appeals to the mutual benefits of cooperation are far more common than explicit threats, while requests for and promises of trustworthy behavior are roughly as frequent as explicit threats. The frequency of appeals to mutual benefits is flat across all ten rounds, but the use of promise of and requests for trustworthy behavior grows over time in a manner similar to growth of the use of threats.

The remaining three panels of Figure 7 illustrate the effectiveness of these three categories in the PChat treatment, showing the likelihood of cheating subject to whether one of the categories was received or sent. Either receiving or sending appeals to the mutual benefits of cooperation leads to a reduction in cheating, particularly for the middle rounds, but this positive effect is never as large as the gain from explicit threats and largely vanishes by the end of the experiment. Receiving a promise of trustworthy behavior modestly reduces the likelihood of cheating, an effect which vanishes in the later rounds, while receiving a request for trustworthy actually leads to slightly more cheating. Sending either type of message about trustworthy behavior is associated with less cheating, but the effect is weak and inconsistent. Overall, it does not appear that the use of communication relating to other-regarding behavior is a strong explanation for the improvement in cooperation for the late rounds of the PChat treatment.

[Figure 8 here]

Figure 8 reports the same information for the RChat treatment the Figure 7 showed for the PChat treatment. Looking at the upper left panel, usage of these three categories in the RChat treatment is roughly the same as in the PChat treatment. However, use of all three categories nearly vanishes by the end of the experiment. Looking at the other three panels, there is no compelling evidence that receiving or sending these types of message is having a great impact on the

likelihood of cheating. The data suggests that appeals to the mutual benefits of cooperation, promises of trustworthy behavior, and requests for trustworthy behavior have little effect in the RChat treatment and are therefore abandoned over time.

d. Regression Analysis

Figures 6 – 8 suggest what types of pre-play communication lead to Period 1 collusion, but there are problems with reading too much into these raw statistics. Most dialogues include many codes and there is correlation between the coding of various categories. For example, codes for the three categories related to trust (indicating you should be trusted, expressing distrust, and requesting trustworthy behavior) are positively correlated since these most often occur as part of a general conversation about trust. Furthermore, Figures 6 – 8 include data from all observations including those where subjects agree on something other than mutual play of High for Period 1, and the types of messages sent are correlated with the type of agreement reached.[35] The preceding suggests the need for regression analysis that measures the effect of a message on the likelihood of cheating controlling for what other messages are sent and what type of agreement was reached.

Table 3 reports on a set of such regressions in which the dependent variable is a dummy for whether the subject cheated on an agreement (i.e. choose a price lower than the agreed upon price). The few observations where an agreement was not reached are dropped from the data set. Since cheating is a binary variable, all three regressions use a probit specification. Standard errors are corrected for clustering at the subject level.

The independent variables of greatest interest are measures of whether a comment from a specific category was *received* by the individual or *sent* by the individual in the dialogue prior to Period 1. The regressions incorporate all categories that were coded in at least 10% of dialogues (averaging across coders) for the treatment in question, including several categories that have not previously been discussed. There are three exceptions to this rule. Dummies for the agreed upon price are included as independent variables, so categories that relate to reaching a pricing agreement, such as proposing play of High in Period 1, are excluded. The category for agreeing to a punishment scheme is excluded since receiving a message of this type is highly correlated with sending a threat, making the standard errors imprecise due to colinearity.[36] Finally, because we are particularly interested in the effect of threats, the categories for implicit and explicit threats were included even when the 10% threshold was not reached.

[35] For example, there is strong positive correlation in the PChat treatment between observing an explicit threat and agreeing to choose High in Period 1.

[36] About two-thirds of subjects receiving a message explicitly proposing a punishment send a message agreeing. No other pair of independent variables has such a high level of correlation.

$$Cheat_{it} = \alpha + \sum_{RdCat=2}^{5} \left(\beta_{RdCat} d_{RdCat}\right) + \sum_{\substack{Cat \in \\ included\ categories}} \left(\rho_{Cat} Received_{it}^{Cat} + \sigma_{Cat} Send_{it}^{Cat}\right) + \lambda_H d_{Medium} + \lambda_L d_{Low} + \varphi Ave_Per1_i + \gamma Ave_Opponent_Per1_{it} + \varepsilon_{it} \quad (7)$$

Equation 7 shows the full specification being estimated, with $Cheat_{it}$ giving the latent variable, where i indexes subjects and t indexes rounds. $Received_{it}^{Cat}$ is the variable for subject i receiving a message coded under category "Cat" in Round t and $Sent_{it}^{Cat}$ is the variable for subject i sending a message coded under category "Cat" in Round t. These variables are averages across the two coders and therefore have three possible values of 0, ½, and 1. The variables d_{Medium} and d_{Low} are dummies for agreements to play Medium or Low, respectively, in Period 1. Agreement to play High is the excluded category. All of the regressions include dummies for each two round block to control for pure experience effects (d_{RdCat}).

Beyond clustering at the subject level, we control for individual effects by including controls for a player's average Period 1 price in Rounds 1 – 10 (Ave_Per1_i) as well as the average Period 1 price in Rounds 1 – 10 of their opponent in Period block t ($Ave_Opponent_Per1_{it}$). Both variables are included to address potential omitted variable bias. Ave_Per1_i is included in the regression because an association between some types of communication and the absence of cheating need not imply a causal relationship, if individuals who have a propensity to cooperate also tend to make certain types of comments. The opponent's behavior in Rounds 1 – 10 is not observed by a subject but is included in the regression to control for omitted variable bias due to imperfections in the coding scheme. The idea is that individuals who are inherently more cooperative, as measured by this variable, may also communicate in subtly different ways not captured by the coding scheme. If these types of statements correlate with the use of any of the included code categories, $Ave_Opponent_Per1_{it}$ can absorb any such effects.

A third source of omitted variable bias can arise because any dialogue between two players is inherently interactive. It can therefore appear that receipt of some particular comment is causing cooperative action, when the effect is really driven by an earlier message from the recipient to which the sender is responding. Including controls for sent messages eliminates this source of bias as the impact of sent messages is directly accounted for. Unlike the variables for what types of communication were received, we cannot interpret the variables for what types of communication were sent as identifying a causal relationship with cheating since it is possible that sent messages and cheating are jointly determined by a common unobserved individual effect. For example, making promises to behave in a trustworthy fashion is associated with not cheating, but both tendencies could be driven by an individual tendency to be trustworthy. The coefficients on sent messages in the regressions reported in Table 3 therefore have to be interpreted with care.[37]

[37] In our discussion of the regression results we argue that at least some of the effects of sent messages cannot be attributed to individual effects, but this relies on regression analysis beyond the specifications in Table 3.

[Insert Table 3]

Models 1 and 2 (the first two columns on Table 3) use data from the PChat treatment.[39] Model 1 does not include controls for what messages were sent by the other player while Model 2 includes these variables. The inclusion of the category variables for sent statements somewhat affects the estimated effects of receiving messages, particularly for messages expressing distrust or appealing for trustworthy behavior. However, the qualitative conclusions are unaffected by including controls for sent messages. Only receipt of an explicit threat has a significant effect on the likelihood of cheating. The estimated marginal effect is large, with a 25% reduction in the probability of cheating. Receiving a message that appeals to mutual payoffs to justify mutual cooperation barely misses statistical significant at the 10% level. While not statistically significant, receiving an implicit threat or an appeal for trustworthy behavior leads to *more* cheating with fairly large marginal effects in both cases (16% and 11% increases in cheating respectively). The regression includes four categories of messages that relate to trust or mutual payoffs. We have looked at joint significance of receiving messages in these four categories, as well as subsets of these categories, but find no jointly significant combinations. Looking at interactions between different categories of messages also yields little in the way of new insights.

Looking at what messages were sent, subjects who send an explicit threat are significantly less likely to cheat. Subjects who promise to be trustworthy were also significantly less likely to cheat on an agreement. To clarify whether these two effects are due to uncontrolled individual effects, for each of these types of message we created a new variable measuring whether the subject had *ever* sent a message of that type. Because of averaging across coders, this is a trinary variable taking on values of 0, ½, and 1. We then reran Model 2 with these two new variables included. If the effects of sending an explicit threat or promising to be trustworthy reflect uncontrolled individual effects, the new variables should be statistically significant and the variables measuring when these messages are actually sent should not be statistically significant. In other words, what should matter is whether a subject is the type of individual who would send one of these messages, not whether the message he sent the message in the current round. The estimated parameters for the two new variables are small and nowhere close to statistical significance.[40] The parameter estimates for sending these types of messages are virtually unchanged in magnitude and both remain statistically significant at the 5% level.[41] This suggests

[39] The parameter estimates for the agreement dummies, round block dummies, and controls for early Period 1 prices are not of direct interest and are therefore suppressed in Table 3 to save space.The only statistically significant effect from the controls for early Period 1 prices is that subjects in the PChat treatment are significantly more likely to cheat when their opponent was more cooperative in Rounds 1 – 10. Exclusion of these two variables has no qualitative effect on our conclusions about the impact of sending or receiving messages in either of the chat treatments.

[40] The parameter estimates for the two new variables are -.180 and -.088 with standard errors of .342 and .280 respectively.

[41] The parameter estimates are -1.070 and -.531 with standard errors of .428 and .254.

that that it is unlikely that individual effects are responsible for the significant effects of sending explicit threats or promises of trustworthy behavior.[42]

The choice of a 10% frequency as the cutoff for including categories in the regression analysis is admittedly arbitrary. We therefore reran Models 2 including all categories that were coded in at least 5% of the dialogues for the treatment in question. This adds five categories for the PChat treatment. The effect of adding these additional categories is minimal. The parameter estimates and standard errors for the previously included categories change little, although the estimate for appeals to mutual payoffs does edge up to statistical significance at the 10% level. Only one of the new categories has a significant effect on the probability of cheating. Oddly, receiving an appeal for mutual trust (i.e. "let's not screw each other over, ok") leads to significantly more cheating. We are inclined to be cautious about this result since the effect is only significant at the 10% level and this is the least frequently observed category in the regression.

Regularity 4: In the PChat treatment, explicit threats have a large and persistent effect of reducing cheating on agreements. Justifying cooperation by appeals to mutual payoffs also has a modest effect of reducing cheating.

As we have seen, Period 1 communication in the RChat treatment has very different content and impact Period 1 communication in the PChat Treatment. This point is reinforced by the results of Model 3 which estimates the full model, including variables for sent messages, using data from the RChat treatment. Notice that Models 2 and 3 include different communication categories since the 10% rule is applied by treatment. Consistent with our observations from Figure 6, receiving or sending an implicit threat lead to significantly less cheating in the RChat treatment. While the parameter estimate looks much bigger than the estimate for explicit threats in the PChat treatment, the estimated marginal effects on the probability of cheating are about the same (27% vs. 25%). The estimated effect of explicit threats on cheating is positive, but small and not close to statistically significant.[43] Sending a message reporting having been cheated earlier is marginally significant as a predictor of less cheating.

To check whether the significant effects of sent messages can be attributed to uncontrolled individual effects, we modified Model 3 to include variables measuring whether a subject was *ever* coded for sending an implicit threat, an explicit threat, or a message reporting having been cheated earlier. All three parameter estimates are small and fail to approach statistical signficance.[44] With the inclusion of these three variables, the estimated effect of sending a message reporting having been cheated previously is reduced by more than 50% and is no longer statistically significant.

[42] As an alternative, we have added variables measuring the number of times each type of message was sent. This yields similar results.
[43] There are only 8 observations in the RChat treatment where one or more coders coded an explicit threat, so it isn't surprising that the estimate is imprecise. This category failed to make the 10% threshold but was included in the regression because of our particular interest in threats.
[44] The parameter estimates are .295, .068, and -.183 with standard errors of .368, .571, and .284.

We reran Model 3 including all categories that were coded in at least 5% of the dialogues for the treatment in question. This adds three categories for the RChat treatment. Critically, the categories for indicating that you should be trusted and appeals for trustworthy behavior are added. The effect of the additional categories is minimal, with the only statistically significant effect coming from sending a promise to be trustworthy. As in the PChat treatment, subjects who promise to be trustworthy are less likely to cheat with the parameter being statistically signficant at the 1% level. The estimated effect of receiving an implicit threat is largely unaffected by the added categories, but the effect of reporting having been cheated in earlier rounds is no longer statistically significant.

Regularity 5: In the RChat treatment, implicit threats have a large effect on reducing cheating.

f. Learning Effects of Chat

Thus far we have focused on the immediate effects of various types of communication on actions taken. However, there are also two ways in which subjects may learn from chat. First, they may learn about effective statements from others. Second, they may learn to respond differently to received messages, potentially making cheating less likely in the future because beliefs about others' actions (e.g. belief in social norms) are changed.

We do observe that receiving a message affects the likelihood that a subject will use that type of message in the future. For example, players who did not use explicit threats in the previous round are more than three times as likely to use an explicit threat if they received an explicit threat. The growth of this and other common types of chat, as documented in Figures 6 and 7, largely reflects delayed similar effect that individuals adopt types of messages that they have seen opponents use.

[Figure 10 here]

We also observe an impact of receiving certain messages today on future cheating whether or not the subject sends such messages herself in the future. Figure 10 shows these effects. We focus in our discussion on data from the PChat treatment since, unlike RChat, this treatment has obvious dynamics in the frequency of cheating. For each of the five common categories of chat discussed in Sections 4.b and 4.c, Figure 10 displays the likelihood of cheating in Round t + 1 subject to whether or not a message from this category was received in Round t. For all categories except requests for trustworthy behavior, receipt of a message in Round t is associated with less cheating in Round t + 1. Explicit threats have the largest immediate effect on cheating but have a relatively small lagged effect.

Figure 10 does not separate delayed effects of receiving a message that are due to adopting that message from effects that are due to changes in beliefs. To measure the lagged effect of receiving messages more formally, we reran Model 2 from Table 3 with additional variables controlling for the previous round's outcome – the messages sent and received, whether the player cheated on an agreement, and whether their opponent cheated. Observations from Round 11 are dropped to allow for the use of lagged variables. Receiving an appeal to mutual payoffs in Round t leads to

significantly less cheating in Round t + 1.[45] This is the only category for which receipt of a message in Round t significantly affects the probability of cheating in Round t + 1. Receiving an explicit threat in the previous round makes subjects more likely to use an explicit threat in the current round, but once we control for this they are no less likely to cheat in the current round.

The lagged effect of appeals to mutual benefits is consistent with our conjecture that this type of message primes a social norm that joint payoffs are important. More generally, chat affects the likelihood of cheating through channels beyond the immediate effect. We do not claim to have catalogued all the indirect effects of chat that exist in our data, but merely point out that the effects of chat can be subtle and can linger beyond the moment in which a message is sent.

e. Chat and the Incentives to Cheat

In this section we analyze whether we can resolve the puzzle posed by Figure 5 that subjects cheat less than they should given the average incentives monetary incentives faced in both the PChat and RChat treatments. We explore the possibility that subjects face very different cheating incentives depending on the type of chat messages they receive. Models 2 and 3 from the regression analysis have identified several message types that are associated with lower rates of cheating when received. We would therefore expect subjects that receive messages that are effective in reducing cheating may face low incentives to cheat while subjects that receive messages that are ineffective in reducing cheating may face high incentives to cheat. .

For each of the chat categories that we identified in the regressions, Figure 9 displays the effect of receiving or sending this type of message on the incentives to cheat. The top panel shows data from the PChat treatment and the bottom panel shows data from the RChat treatment. Data is only included from observations where the subjects agreed to play High in Period 1. For each case the average total payoffs (sum of Period 1 and 2 payoffs) are plotted conditional on whether or not the player cheated on the agreement. Hollow bars indicate cells with less than five observations. Two cells had no observations (and therefore are missing). As a point of comparison we also plot the total payoffs for all observations with agreements to play High, conditioned on whether or not cheating takes place.

[Figure 9]

Figure 9 must be interpreted with caution given the large number of cells with sparse observations. None the less, several points seem clear. With one exception we see that the types of communication that reduce cheating also reduce the incentives to cheat. For receipt of explicit threats in the PChat treatment or implicit threats in the RChat treatment, this clearly works to a

[45] The parameter estimate is -.414 with a standard error of .200. This is statistically significant at the 5% level. We have tried a variety of other specifications including only subsets of the lagged variables. The effect of lagged appeals to mutual benefits is robust. This is the only lagged category which always has a statistically significant effect. In some specifications, lagged receipt of a promise of trustworthy behavior reduces cheating or lagged receipt of a request for trustworthy behavior increases cheating. Comparing across specifications, these two categories generally have a large estimated effect (similar in magnitude to the estimated effect for an appeal to mutual payoffs) but the estimates lack precision.

large extent through increasing the *payoff from not cheating*. The subjects sending such threats almost never cheat. This implies that subjects receiving these threats have less of an incentive to cheat themselves. Sending threats also improves the payoffs from not cheating – there are too few observations to say what the effect following cheating might be. Sending a promise of trustworthy behavior in the PChat treatment also improves the incentives to not cheat, but more by reducing the payoff from cheating rather than raising the payoff of not cheating. Subjects who promise to cheat and then renege are punished harshly. In all of these cases it appears that the reduced cheating incentives are well explained by the a change in the incentives that a firm faces after receiving or sending a specific category of messages. The one exception is that of receiving a message appealing to mutual payoffs. In this case the incentive to cheat does not become smaller but, possibly, slightly increases.[47] Nevertheless, the rate of cheating is lower for subjects that appeal to mutual benefits.

These results suggest three channels by which cheating incentives are reduced. One is based on purely financial incentives generated is through coordinating on contignent behavior, consistent with our models of collusion. This is explicit in the case where explicit threats are being made and, as will become clearer below, implied in the case of implicit threats in the RChat treatment. A second way in which chat and financial incentives seem to interact is through chat providing a self-commitment effect: Sending a promise of trustworthy behavior helps maintain the collusive equilibrium by putting the sender in greater danger of punishment if he cheats. Finally, chat may aid collusion by changing the psychological benefits of not cheating. This is consistent with the reduced cheating that follows from receiving an appeal to the mutual benefits of collusion in the PChat treatment. The incentives to cheat are higher than normal when such messages are received, which suggests that reduced cheating must be due to something else. We conjecture that such messages are priming a social norm that mutual benefits are important and hence making subjects less willing to cheat.

These results resolve the puzzle for PChat that was generated by Figure 5. The reason why individuals on average have incentives to deviate is reflected in the high cheating probability for those individuals who do not use effective chat, while using effective chat leads to low cheating probabilities, which are generally consistent with the monetary incentives. However, this observation is not true for RChat. We now show that the puzzle for RChat can be resolved by analyzing second period chat.

g. The Effect of Renegotiation

The effects of pre-play communication documented above cannot explain why Period 1 collusion is so consistently high in the RChat treatment. Implicit threats are quite effective in this treatment, but do not become common until the later rounds while collusion is high throughout. Even in observations where no implicit threat is observed, cheating is relatively rare in the RChat treatment across all rounds (see Figure 6). Likewise, it is puzzling that implicit threats are so effective in the RChat treatment given that they are not especially effective in the PChat treatment (and indeed the theory of cheap talk gives no reason why non-specific threats should be useful).It

[47] This remains true if we limit ourselves to observations from Rounds 16 – 20.

is hard to explain the high levels of cooperation by appealing to some other category of pre-play communication since all other types of communication are less frequent than in the PChat treatment (see Table 2) and lack the same strong effect on cheating that implicit threats have.

The preceding observations suggest that the explanation of high collusion in the RChat treatment must lie in the content of communication between the Period 1 and 2 decisions. The theory of renegotiation suggests that this chat should reduce collusion in Period 1 by making contingent punishments non-credible, but in actuality we will see that Period 2 chat plays a central role in generating high collusion in the RChat treatment.

Consistent with the theory of renegotiation, attempts at renegotiation following unilateral cheating are frequent and reasonably successful. For games in which the players agreed on High for Period 1 and then one player cheated, mutual play of High in Period 2 is suggested by one of the players in 91% of the dialogues between Periods 1 and 2. When such suggestions occur, 75% of the players actually choose High in Period 2. This compares with 33% for the (admittedly infrequent) cases where play of High is not suggested. Not surprisingly, the impact of attempts at renegotiation is larger for cheaters than those for who were cheated and larger if the suggestion comes from the player who was cheated. The frequency and success of renegotiation causes the financial punishment of cheating to be relatively low in the RChat treatment, as shown in Figure 4. It should, however, be noted that some financial punishment still occurs.

The following dialogue gives a sense of why cheating is infrequent in the RChat treatment even though the financial punishment of cheating is weak. These two players had agreed to both choose High in Period 1 with minimal discussion. Player 1 followed this agreement by choosing High, but Player 2 cheated and chose Medium.

> 1 YOU MEANIE!!!!!!!!!!!
> 2 its in both our interests to choose [Period 2 High]
> 1 no
> 1 i wont help you
> 2 but you hurt yourself in the process
> 1 you already hurt me :(

Player 1 has an immediate negative response to Player 2's cheating. Player 2 responds by trying to negotiate Period 2 cooperation, first pointing to mutual benefits from coordinating on High and then pointing out that Player 2 hurts himself by not agreeing to mutual play of High. As Player 1's response makes clear, he is more interested in punishing Player 2's misdeeds than in making the highest possible payoff. Player 2 apparently did not get the hint, choosing High in Period 2 while Player 1 chose Low.

The type of negative verbal response to cheating that Player 1 makes in the preceding example is quite common. In cases where one of the players cheated and the other did not, 73% of the

players who got cheated admonished the other player for either cheating or lying.[49] The messages were often quite emotional, could be very personal in nature, and frequently ignored the instructions to avoid cursing. Some additional examples include "good job, [expletive deleted]," "you are a bad person . . . i hope somone [expletive deleted] you over as well", and (our favorite) "you know, they shoot you for that in Texas." Messages of this type are best interpreted as strong non-pecuniary punishments in the spirit of Noussair *et al*. Subjects who were cheated and sent a verbal punishment were more likely to engage in conventional punishment as well – 42% of subjects who were cheated and admonished their opponents also chose a price other than High, compared with only 22% of subjects who were cheated and did not admonish their opponent. This runs contrary to the results of Xiao and Houser (2005) who argue that verbal punishment is a substitute for conventional punishment rather than a complement.

The possibility of verbal punishment provides a cheap way to punish cheating. If subjects dislike being told off, the result should be less cheating than when only conventional punishment (choosing low prices in Period 2) is possible. Indeed, receiving a verbal punishment deters future cheating. Subjects who unilaterally cheat on an agreement to play High in Period 1 and get admonished for cheating only cheat in 40% of the observations for the following round, compared with 62% for cheaters who do not get admonished. The proportion of future cheating drops even more, to 27%, when a cheater is also admonished for lying. In contrast, cheaters who receive conventional punishment are slightly more likely to cheat in the following round (50%) than those who do not (46%).[50] Admonishment for cheating and lying appears to be a particularly powerful form of punishment. We speculate that this is due to its clarity. If a subject is punished for cheating by a lower Period 2 price, they may fail to grasp that this was due to their cheating. No misunderstanding is possible with verbal punishment, since the cheater is explicitly told that his first period actions were "ridiculous," "stupid," or "retarded" among other things.

Turning to messages sent (rather than received) by cheaters, they often either apologized or tried to justify their actions. The dialogue below gives a good example of this. Prior to Period 1 the two players had quickly agreed to both choose High. Player 1 deviated by choosing Medium.

> 1 sorry about that.....
> 2 ...and that is why I'm cynical
> 1 I assumed you would have screwed me over like the last person I had
> 2 the dark side of capitalism
> 1 well, I can assure you that I am going [Period 2 High] this time
> 1 very true
> 2 there's no reason not to
> 2 ok, selecting [Period 2 High]

[49] To be precise, at least one of the two raters coded this category in 73% of the observations. The number reported in Table 2 differs slightly since this table reports an average across the two coders.
[50] Subjects who receive conventional punishment are more likely to cheat in the following round whether they receive admonishment or not.

In this case both players selected High for Period 2. Apologies and justifications are not necessarily sincere – one memorable subject cheated for Rounds 12 – 20 and justified it every single time by saying their opponent had done the same to them in the previous round. In fact, they did not face a single person who cheated! Surprisingly, apologies and justifications are somewhat effective. Subjects who were cheated and receive an apology are more likely to choose High in Period 2 (71% vs. 55%) as are those who receive a justification for cheating (74% vs. 53%). It is hard to know whether subjects who were cheated truly believed these excuses (e.g. "i was playing with the contrast") or simply looked for a justification to not go through with a costly punishment.

Regularity 6: Unilateral cheating in the RChat treatment is frequently admonished in strong terms prior to Period 2 play. Receiving admonishment leads to less future cheating, suggesting that it is a powerful punishment.

The content analysis of Period 2 chat provides a clear answer for our questions about the RChat treatment. The use of verbal punishments provides a strong non-financial reason to not cheat on agreements, leading to stable collusion. The analysis of Period 2 chat also suggests why implicit threats are so effective in RChat. Implicit threats indicate the attitude of a subject toward cheating and may signal a strong verbal reaction in the future. Implicit threats therefore help recipients anticipate verbal punishment.[51] The logic underlying the theory of renegotiation works just fine in our data, but simply is overwhelmed by the impact of verbal punishments.

5. Conclusions

The primary purpose of the experimental research laid out above was to develop a rich data set that enabled us to study how communication facillitates the development of stable collusion. Our results indicate that allowing a rich message space leads to the development of persistent collusion. As predicted by the theory of collusion, the use of explicit threat to punish deviation from collusive agreements is the most effect type of message for promoting collusion when only pre-game communication is allowed. Other types of messages, particularly those relating to other-regarding preferences, have some role to play as well. This is especially true in the RChat treatment where the strong incentives created by verbal punishments lead to extremely high and stable levels of collusion.

Our results on renegotiation suggest that social context plays an important and theoretically underappreciated role in collusion. A natural question to ask is whether results that rely so heavily on an emotional reaction from subjects will extend to field settings. The cheap criticism of laboratory experiments is that the usual undergraduate subjects are "different" from businesspeople. This may be true at some level, but begs an obvious question. Business people are human beings who typically went to college and presumably have a functioning emotional capability. To claim there will be a difference requires a theory of why such differences might exist. Many important differences between the lab and the field are likely to be driven by the social context rather than any inherent difference between businesspeople and the rest of the population. In particular, almost all business decisions are made in a group context and within an

[51] We cannot test whether subjects who make implicit threats are more likely to use verbal punishment because there are so few observations where a subject makes an implicit threat and is cheated. There are no examples where a subject explicitly threatened a future verbal punishment.

organizational hierarchy. Individuals in these settings may feel more responsible to other people within their corporation than to any person in another firm. In other words, defining identity of the relevant "other" is an important and under-appreciated aspect of applying other-regarding preferences to field settings. A central topic for future research is exploring how our results on renegotiation are affected by changing the organizational structure within which subjects operate.

Our experimental design also raises serious methodological questions. Using a simple game like the TSBG has obvious advantages, but could lead to different outcomes than would be observed in an indefinitely repeated game. The rich communication allowed by our chat treatments allows us to observe the impact of a wide variety of types of messages, but also greatly increases the complexity of analyzing the effects of communication. Limited message space experiments ease this complexity, but possibly at a cost of changing how messages are used or eliminating vital components of communication. In future work we plan to explore the implications of using a finite game versus an indefinitely repeated game and the implications of using limited versus rich message spaces to study communication and collusion.

Appendix 1: Statistical Analysis of Treatment Effects

To confirm the statistical significance of these treatment effects, Table 2 shows the results of two regressions. For both regressions, the dataset consists of all individual choices from Rounds 11 – 20. The dependent variable is the Period 1 choice. Given that the available choices are naturally ordered categories, we use an ordered probit specification (0 = Low, 1 = Medium, 2 = High). The cutpoints between categories are not reported since these are not of direct interest. All standard errors are corrected for clustering at the subject level.

The two models are designed to answer different questions about the data. Model 1 looks for differences between treatments. The equation being estimated for the latent variable is shown below as (5). The dependent variable is the Period 1 price for subject i in Round t (P_{it}). To control for changes over time, rounds have been broken down into five categories: category 1 (RdCat = 1) is Rounds 11 and 12, category 2 (RdCat = 2) is Rounds 13 and 14, etc. Use of a non-linear specification for time is necessary given the non-monotonic time trend in the Pchat treatment.[52] The variables d_{RdCat} are dummies for the five round categories. The variables d_{P1}, d_{PChat}, and d_{RChat} are dummies for the P1, PChat, and RChat treatments respectively. The interactions between dummies are stacked so we get an estimate of the difference between pairs of treatments in each round category. So, for example, γ_1 estimates the difference between the P1 and N treatments in round category 1 (Rounds 11 – 12), η_1 estimates the difference between the PChat and P1 treatments in round category 1, and ν_1 estimates the difference between the PChat and RChat treatments in round category 1. The variables Ave_Per1_i and Ave_Per2_i give the average Period 1 and Period 2 prices respectively for subject i from Rounds 1 – 10. These averages are calculating setting 0 = Low, 1 = Medium, and 2 = High. These variables are included to better capture the individual effects in the data.

(5)

$$P_{it} = \alpha + \sum_{RdCat=2}^{5} \left(\beta_{RdCat} d_{RdCat}\right) + \varphi Ave_Per1_i + \gamma Aver_Per2_i +$$

$$\sum_{RdCat=1}^{5} \left(\gamma_{RdCat} d_{RdCat} \left(d_{P1} + d_{PChat} + d_{RChat}\right) + \eta_{RdCat} d_{RdCat} \left(d_{PChat} + d_{RChat}\right) + \nu_{RdCat} d_{RdCat} \left(d_{RChat}\right)\right) + \varepsilon_{it}$$

The results of Model 1 strongly support the existence of treatment effects on Period 1 choices. With one exception, all of the pairwise comparisons of treatments for a two round block are statistically significant at the 1% level.[53] The sole exception is that the initial difference (Rounds 11 – 12) between the PChat and RChat treatments is not statistically signficant. The control for average Period 1 price in Rounds 1 – 10 is statistically significant, consistent with the existence of strong individual effects in the data. There is no statistically significant relationship between Period 2 prices in Rounds 1 – 10 and Period 1 prices in Rounds 11 – 20.

[52] Using five round categories rather than round dummies makes reporting results somewhat more manageable and does not affect the conclusions.

[53] Within a round category, this also holds for the pairwise comparisons that are not explicitly made in Model 1 since the treatments have been stacked from lowest to highest Period 1 prices.

Model 2 looks for changes within treatments over time. The most important question this model addresses is whether the dip and recovery in Period 1 prices for the PChat treatment is statistically significant. The equation being estimated is shown in (6). The dependent variable has not changed from Model 1, and Ave_Per1_i and Ave_Per2_i are defined as in Model 1. The variables d_N, d_{P1}, d_{PChat}, and d_{RChat} are dummies for the N, P1, PChat, and RChat treatments respectively. The primary change from Model 1 comes in how the round category dummies are defined. The variable δ_{RdCat} is a dummy for all observation from that round category *and* subsequent rounds. Thus, δ_1 is a dummy for Rounds 11 – 20, δ_2 is a dummy for Rounds 13 – 20, δ_3 is a dummy for Rounds 15 – 20, and so forth. The dummies are set up to estimate the difference in Period 1 play between two consequtive round categories for the same treatment. For example, γ_1 estimates the difference between round category 1 of P1 treatment and the base (round category 1 of the N treatment), γ_2 estimates the difference between round category 1 (Rounds 11 – 12) and round category 2 (Rounds 13 – 14) for the P1 treatment, γ_3 estimates the difference between round category 2 (Rounds 13 – 20) and round category 3 (Rounds 15 – 16) for the P1 treatment, and so on. The η and ν parameters measure equivalent differences for the PChat and RChat treatments.

$$P_{it} = \alpha + \sum_{RdCat=2}^{5}\left(\beta_{RdCat}\delta_{RdCat}d_N\right) + \sum_{RdCat=1}^{5}\left(\gamma_{RdCat}\delta_{RdCat}d_{P1}\right) + \sum_{RdCat=1}^{5}\left(\eta_{RdCat}\delta_{RdCat}d_{PChat}\right) + \sum_{RdCat=1}^{5}\left(\nu_{RdCat}\delta_{RdCat}d_{RChat}\right) + \varphi Ave_Per1_i + \gamma Aver_Per2_i + \varepsilon_{it} \quad \textbf{(6)}$$

Looking at the results for Model 2, the most important issue is whether the u-turn in the PChat treatment is statistically significant. The decline between round categories 1 (Rounds 11 – 12) and 2 (Rounds 13 – 14) is statistically significant at the 5% level and there are smaller (and not statistically significant) declines between round categories 2 and 3 and round categories 3 and 4. The difference between round category 1 (Rounds 11 – 12) and round category 4 (Rounds 17 – 18) is statistically signficant at the 1% level.[54] The downward trend then reverses, as the increase between round categories 4 (Rounds 17 – 18) and 5 (Rounds 19 – 20) is statistically significant at the 1% level. Both the initial decline in Period 1 prices for the PChat treatment and the following recovery are statistically significant changes. If we compare Period 1 prices for round categories 1 and 5 in the PChat treatment, the different is not statistically significant.[55] By the end of the experiment, Period 1 prices in the PChat treatment have returned to the levels reached immediately following the introduction of communication. Turning to the other treatments, decreases in Period 1 prices are statistically significant in all round categories for the P1 treatment. Thus, even though Model 1 indicates there remains a significant difference between the N and P1 treatment for round category 5 (Rounds 19 – 20), we feel confident in stating that play has not converged in the P1 treatment and hence this difference would probably not persist if the experiment ran for more rounds. The RChat treatment shows a weak increase in Period 1

[54] We use a variant on Model 2 to estimate the change between round categories 1 and 4 for the PChat treatment. The parameter estimate for the difference is -.501 with a standard error of .150.
[55] Using a variant on Model 2 to estimate the difference between round categories 1 and 5 for the PChat treatment, the parameter estimate for the difference is .025 with a standard error of .150.

prices. If we compare Period 1 prices for the round categories 1 and 5, the difference is statistically significant at the 10% level.[56]

[56] The parameter estimate for this difference is .495 with a standard error of .270.

Table 1
Summary of Treatments

	NC	P1	PChat	RChat
Number of Sessions	3	3	3	3
Number of Subjects	64	68	64	76
First Period Messages Only		✓		
First Period Chat			✓	✓
First and Second Period Chat				✓

Table 2
Use of Period 1 Messages in Chat Treatments

Message Description	Proportion Observed PChat	Proportion Observed RChat	κ
Period 1 Proposal: Both Play Medium	0.211	0.081	0.802
Period 1 Proposal: Both Play High	0.889	0.975	0.812
Period 2 Proposal: Both Play High	0.936	0.541	0.847
Disagreement with Most Recent Proposal	0.109	0.040	0.343
Agreement with Most Recent Proposal	0.794	0.791	0.840
Implicit Threat to Punish Cheating in Period 2	0.058	0.065	0.532
Explicit Threat to Punish Cheating with Low in Period 2	0.141	0.017	0.755
Agreement with Proposed Punishment (All Punishments)	0.108	0.023	0.451
Request for Proposals	0.077	0.118	0.721
Appeal to Joint Payoffs	0.297	0.157	0.587
Specific Reference to Payoff Table	0.169	0.096	0.694
Promises of Trustworthy Behavior	0.111	0.086	0.624
Expression of Distrust	0.111	0.036	0.448
Appeal for Trustworthy Behavior	0.153	0.068	0.460
Self-Report Having Been Cheated in Earlier Rounds	0.169	0.108	0.812

Use of Period 2 Messages in RChat Treatment

Message Description	Proportion Observed Period 1, Collusion (Both High)	Proportion Observed Period 1, One Cheated	κ
Positive Feedback Following Cooperation	0.283	0.023	0.731
Apology for Cheating	---	0.477	0.778
Rationalizing Cheating	---	0.447	0.671
Admonition for Cheating/Lying	---	0.568	0.623
Period 2 Proposal: Both Play High	0.901	0.894	0.792
Agreement with Most Recent Proposal	0.675	0.394	0.473
Appeal to Joint Payoffs	0.017	0.265	0.366

Table 3: Probit Regressions on Effect of Chat Categories

	Model 1	Model 2	Model 3
Data Set	PChat	PChat	RChat
Number of Observations/Subjects	626/64	626/64	602/76
Received Implicit Threat	.427 (.417)	.482 (426)	-1.227** (.472)
Received Explicit Threat	-.750** (.297)	-.796** (.314)	.365 (.772)
Received Request for Proposals			.453 (.285)
Received Appeal to Mutual Payoffs	-.263 (.176)	-.312 (.192)	-.011 (.314)
Received Specific Reference to Payoff Table	-.061 (.245)	-.136 (.268)	
Received Indicate You Should be Trusted	-.267 (.289)	-.149 (.296)	
Received Expression of Distrust	-.110 (.265)	.180 (.284)	
Received Appeal for Trustworthy Behavior	.138 (.243)	.347 (.256)	
Received Self-Report Being Cheated Earlier	-.323 (.236)	-.170 (.230)	.163 (.264)
Sent Implicit Threat		-.379 (.473)	Perfectly Predicts No Cheating
Sent Explicit Threat		-1.217*** (.375)	Perfectly Predicts No Cheating
Sent Request for Proposals			.226 (.309)
Sent Appeal to Mutual Payoffs		-.238 (.237)	-.180 (.337)
Sent Specific Reference to Payoff Table		.252 (.287)	
Sent Indicate You Should be Trusted		-.567** (.248)	
Sent Expresion of Distrust		.310 (.385)	
Sent Appeal for Trustworthy Behavior		-.463 (.388)	
Sent Self-Report Being Cheated Earlier		-.258 (.241)	-.500* (.303)
Log Likelihood	-345.13	-328.50	-231.37

Note: All regressions include controls for what agreement (if any) was reached, two round blocks, the player's average Period 1 price in Rounds 1 – 10, and their opponent's average Period 1 price in Rounds 1 – 10. Observations where no agreement was reached are dropped. Standard errors are corrected for clustering at the individual level. Three (***), two (**), and one (*) stars indicate statistical significance at the 1%, 5%, and 10% respectively.

Table A.1: Ordered Probit Regressions on Period 1 Choices

Model 1			Model 2		
Variable	Parameter Estimate	Standard Error	Variable	Parameter Estimate	Standard Error
Rounds 13 – 14 (β_2)	-0.477***	0.125	N, Difference Rds 13-14 vs. Rds 11-12 (β_1)	-0.477***	0.125
Rounds 15 – 16 (β_3)	-0.606***	0.204	N, Difference Rds 15-16 vs. Rds 13-14 (β_1)	-0.129	0.211
Rounds 17 – 18 (β_4)	-0.495***	0.191	N, Difference Rds 17-18 vs. Rds 15-16 (β_1)	0.110	0.237
Rounds 19 – 20 (β_5)	-0.904***	0.205	N, Difference Rds 19-20 vs. Rds 17-18 (β_1)	-0.409	0.262
Rds 11 – 12, Difference P1 vs. N (γ_1)	1.856***	0.187	P1 Treatment	1.856***	0.187
Rds 13 – 14, Difference P1 vs. N (γ_2)	1.935***	0.202	P1, Difference Rds 13-14 vs. Rds 11-12 (γ_1)	-0.398***	0.101
Rds 15 – 16, Difference P1 vs. N (γ_3)	1.682***	0.220	P1, Difference Rds 15-16 vs. Rds 13-14 (γ_1)	-0.382***	0.120
Rds 17 – 18, Difference P1 vs. N (γ_4)	1.052***	0.245	P1, Difference Rds 17-18 vs. Rds 15-16 (γ_1)	-0.519***	0.112
Rds 19 – 20, Difference P1 vs. N (γ_5)	1.220***	0.242	P1, Difference Rds 19-20 vs. Rds 17-18 (γ_1)	-0.241**	0.111
Rds 11 – 12, Difference PChat vs. P1 (η_1)	0.933***	0.176	PChat Treatment	2.789***	0.205
Rds 13 – 14, Difference PChat vs. P1 (η_2)	1.071***	0.174	PChat, Difference Rds 13-14 vs. Rds 11-12 (γ_1)	-0.260**	0.122
Rds 15 – 16, Difference PChat vs. P1 (η_3)	1.357***	0.172	PChat, Difference Rds 15-16 vs. Rds 13-14 (γ_1)	-0.097	0.112
Rds 17 – 18, Difference PChat vs. P1 (η_4)	1.731***	0.184	PChat, Difference Rds 17-18 vs. Rds 15-16 (γ_1)	-0.144	0.125
Rds 19 – 20, Difference PChat vs. P1 (η_5)	2.498***	0.182	PChat, Difference Rds 19-20 vs. Rds 17-18 (γ_1)	0.526***	0.128
Rds 11 – 12, Difference RChat vs. PChat (ν_1)	0.264	0.205	RChat Treatment	3.053***	0.226
Rds 13 – 14, Difference RChat vs. PChat (ν_2)	0.569***	0.199	RChat, Difference Rds 13-14 vs. Rds 11-12 (γ_1)	0.045	0.167
Rds 15 – 16, Difference RChat vs. PChat (ν_3)	0.761***	0.197	RChat, Difference Rds 15-16 vs. Rds 13-14 (γ_1)	0.095	0.113
Rds 17 – 18, Difference RChat vs. PChat (ν_4)	0.931***	0.229	RChat, Difference Rds 17-18 vs. Rds 15-16 (γ_1)	0.025	0.192
Rds 19 – 20, Difference RChat vs. PChat (ν_5)	0.734***	0.278	RChat, Difference Rds 19-20 vs. Rds 17-18 (γ_1)	0.329	0.257
Average Period 1 Price (Rds 1 – 10)	0.882***	0.196	Average Period 1 Price (Rds 1 – 10)	0.882***	0.196
Average Period 2 Price (Rds 1 – 10)	0.105	0.114	Average Period 2 Price (Rds 1 – 10)	0.105	0.114

Note: Both regressions contain 2542 observations from 272 subjects. Standard errors are corrected for clustering at the individual level. Three (***), two (**), and one (*) stars indicate statistical significance at the 1%, 5%, and 10% respectively.

Figure 1: Period 1 Play

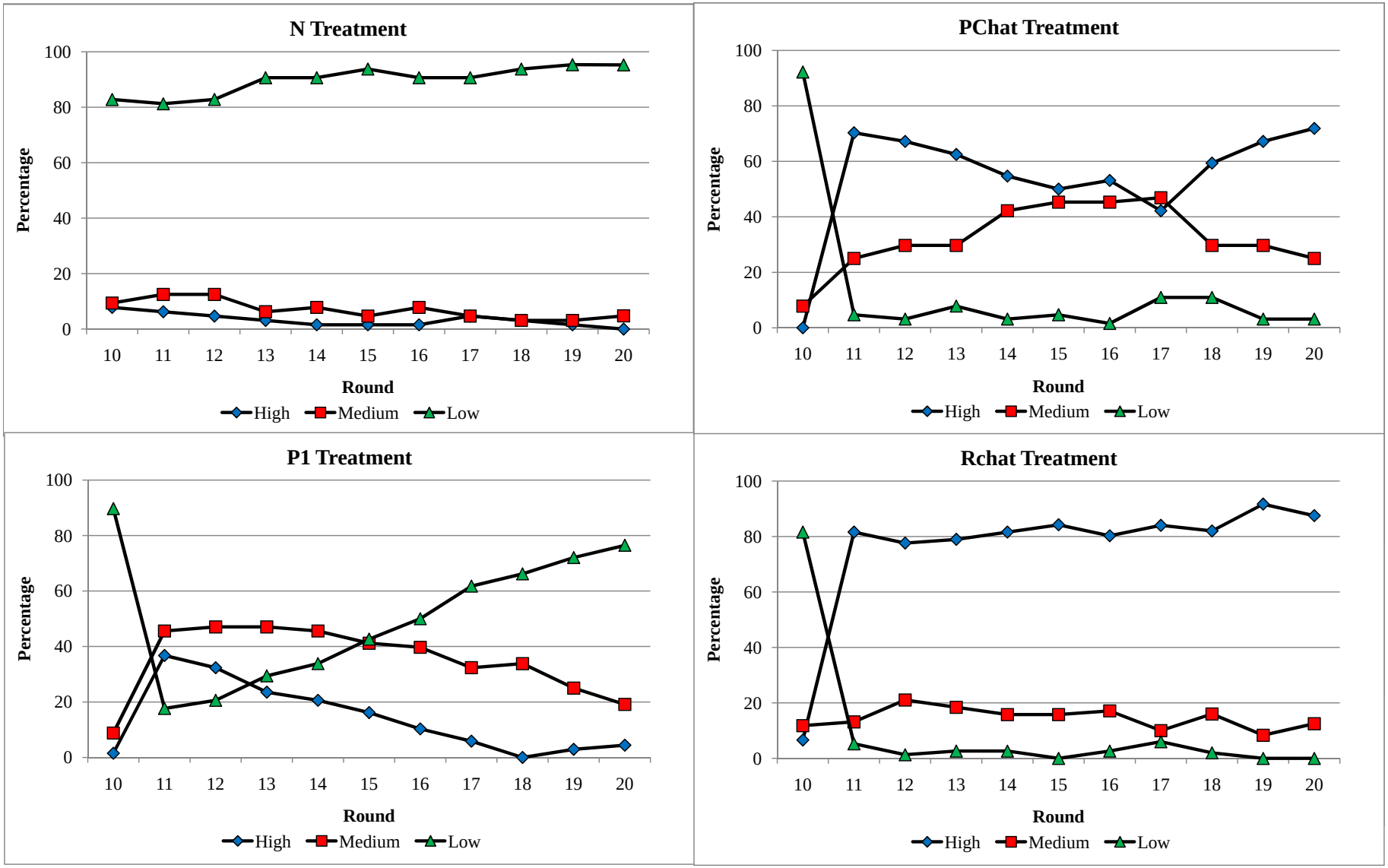


Figure 2: Cumulative Distribution of Agreements on Period 1 Prices

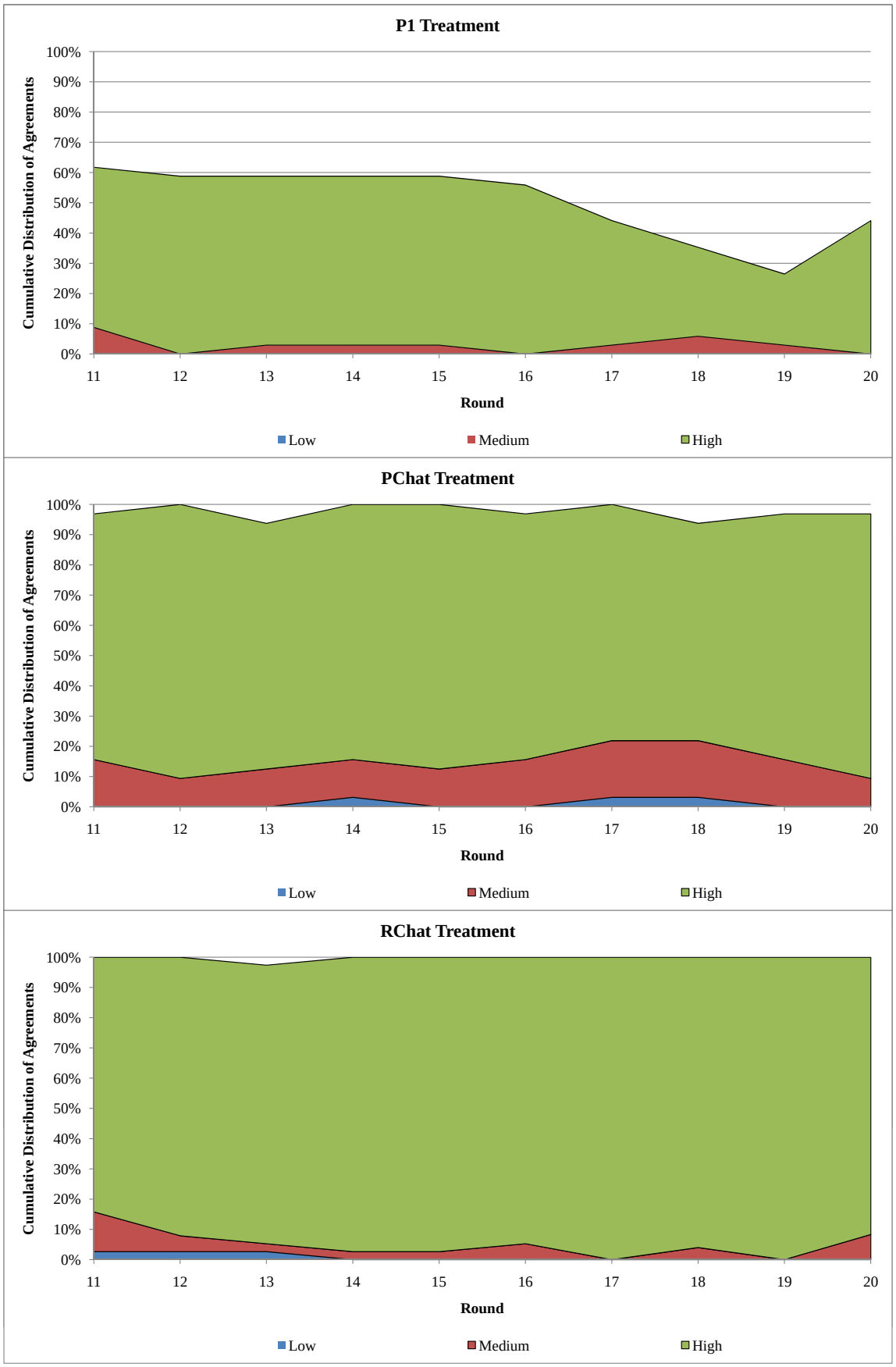


Figure 3: Probability of Cheating on an Agreement to Choose High in Period 1

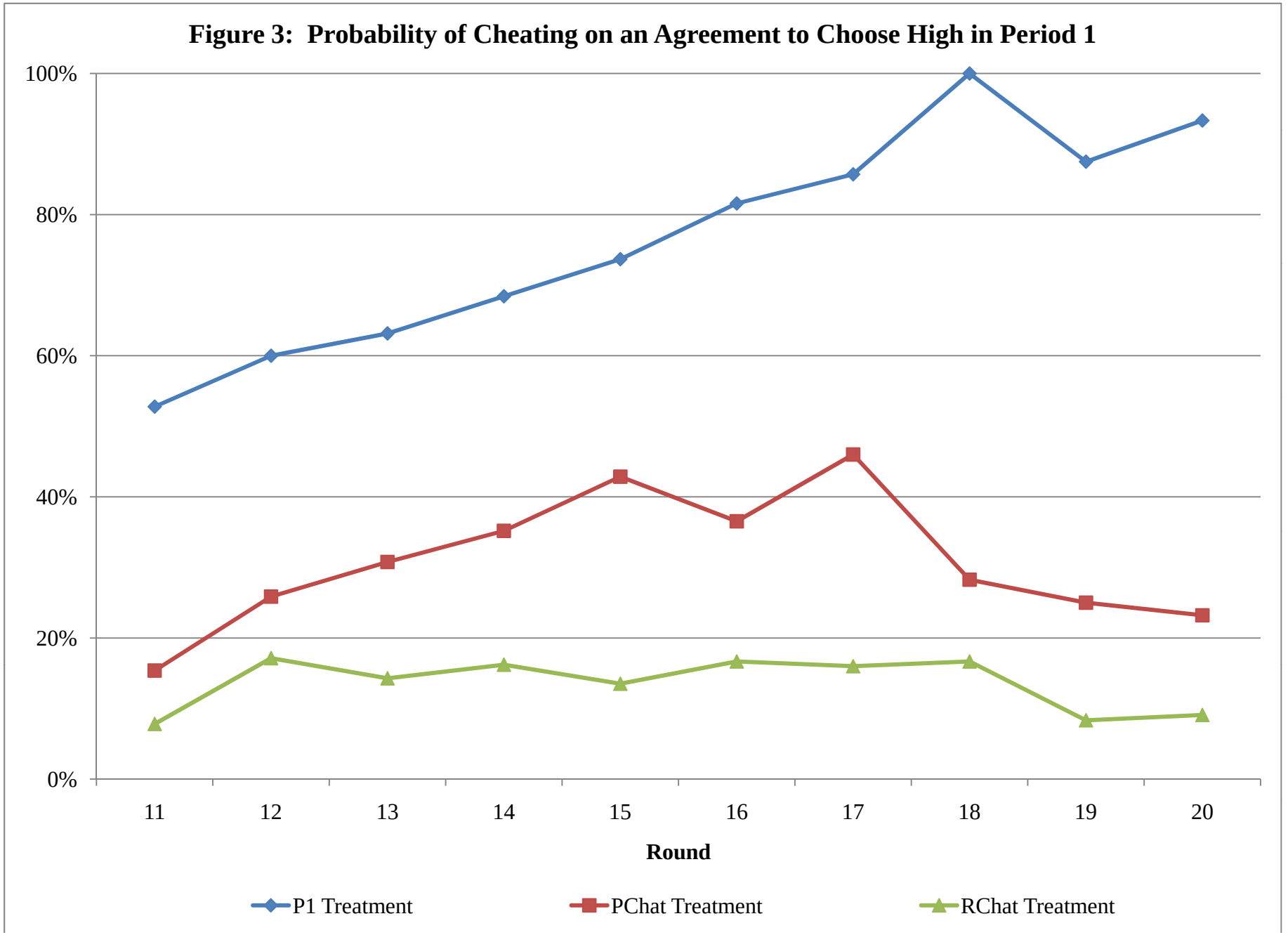


Figure 4: Punishment for Cheating on an Agreement to Play High in Period 1

(Note: Hollow markers are cells with five or fewer observations.)

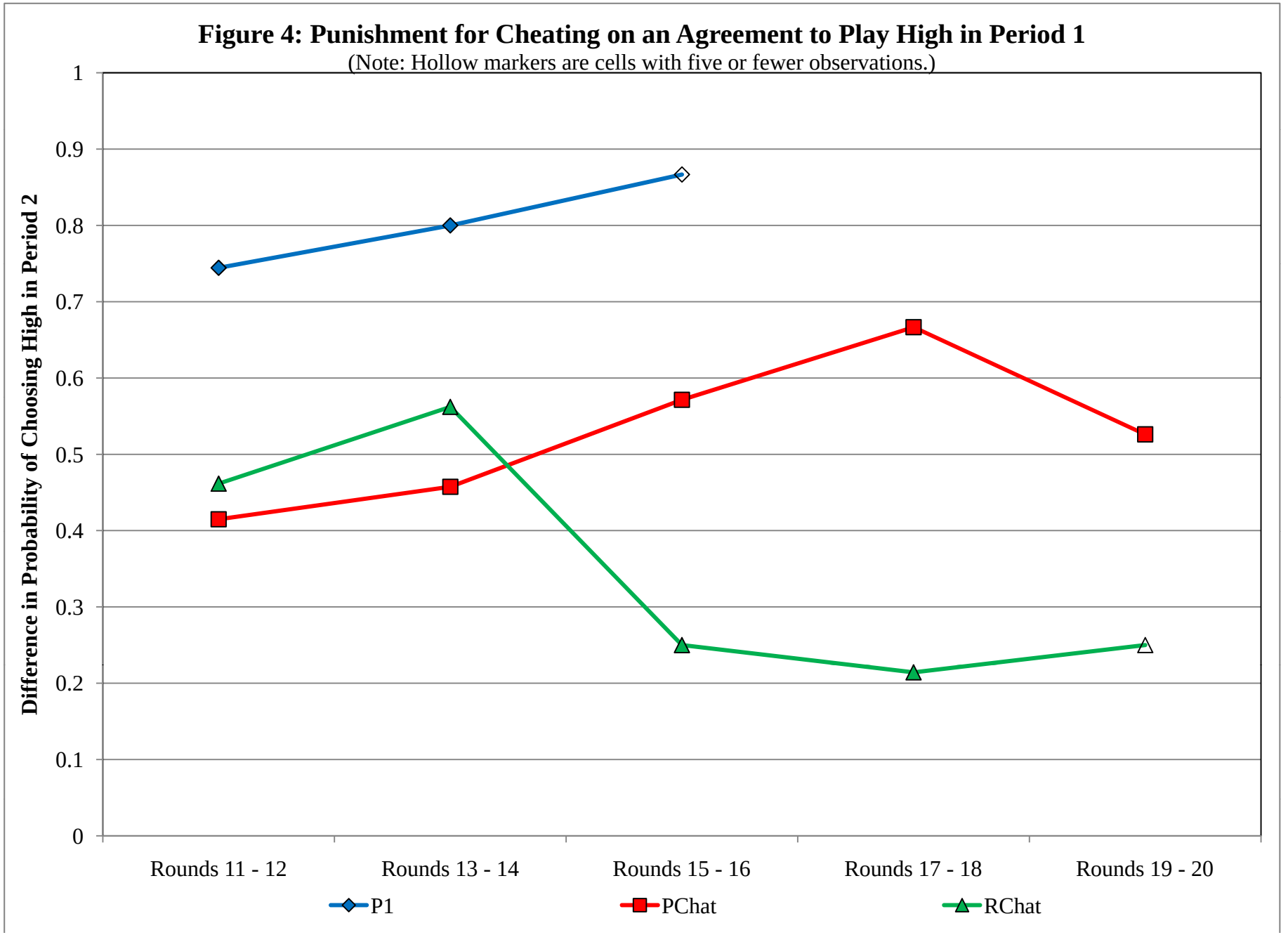


Figure 5: Gains from Cheating on an Agreement to Play High

(Note: Hollow markers are cells with five or fewer observations.)

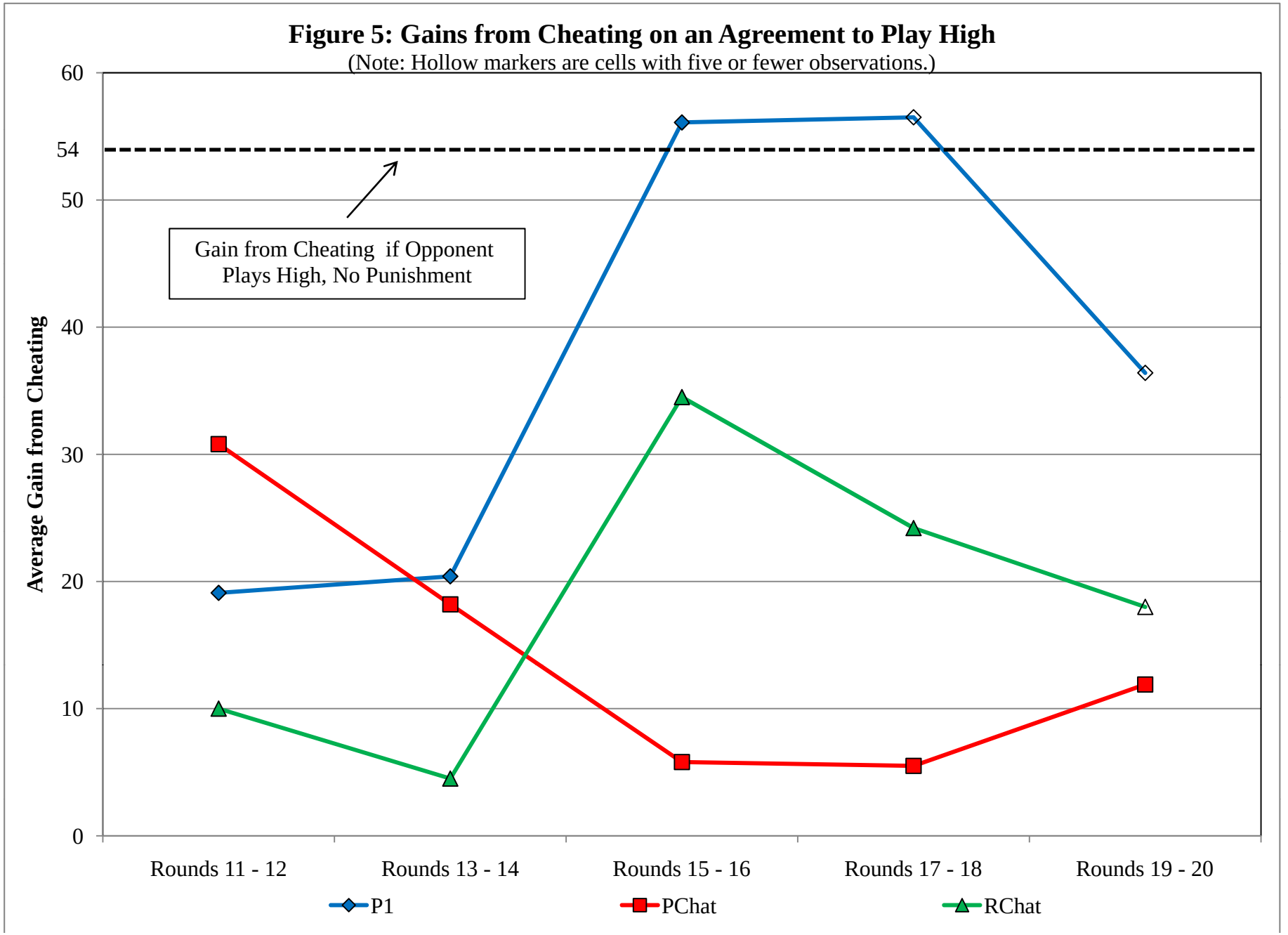
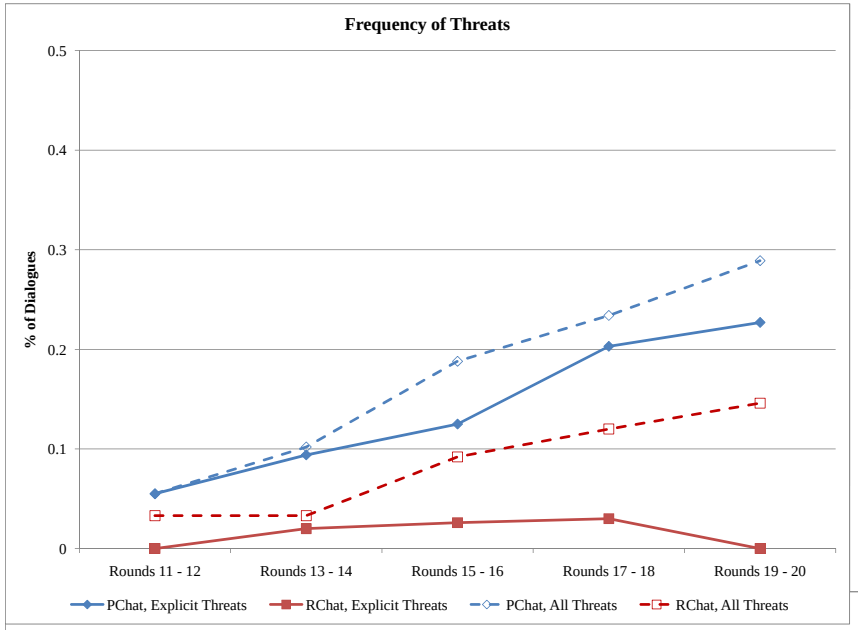
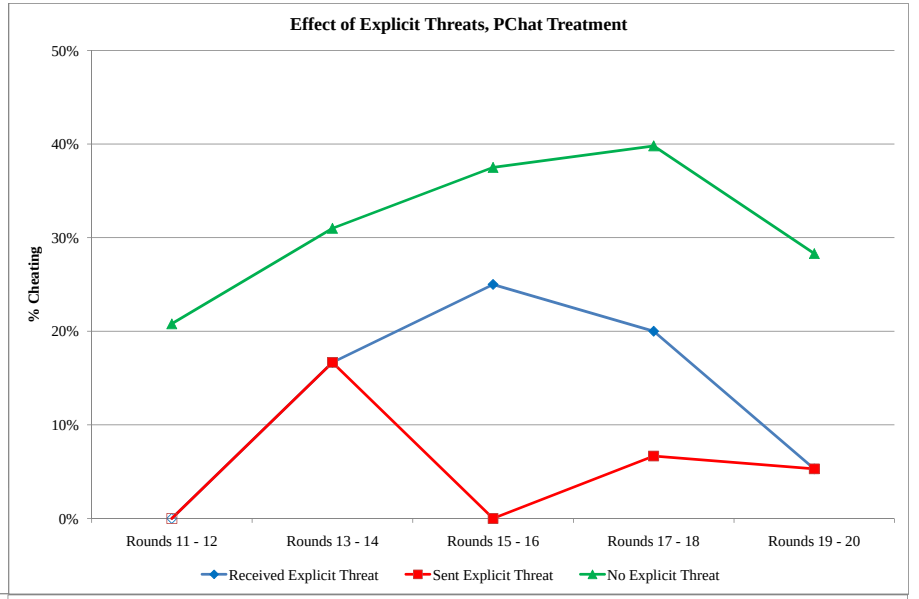
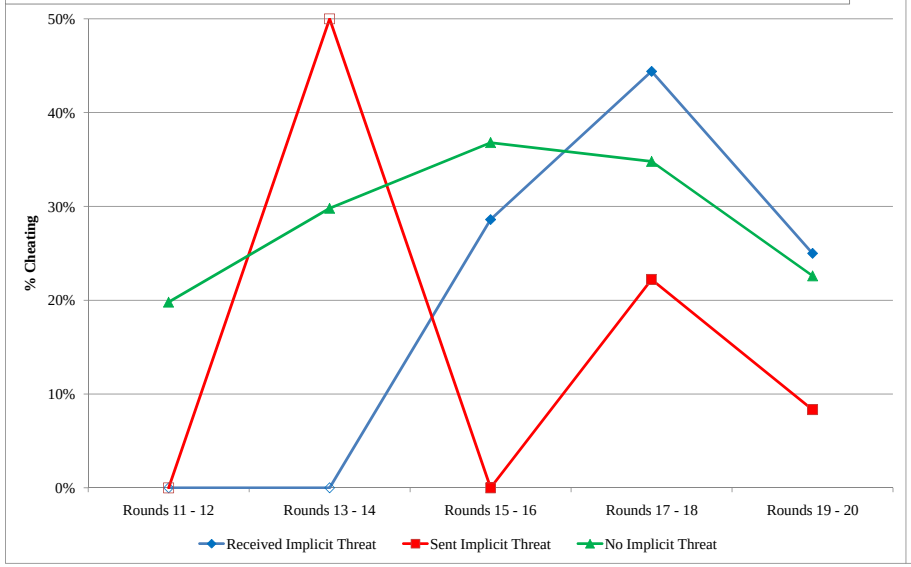


Figure 6: Use and Effect of Threats







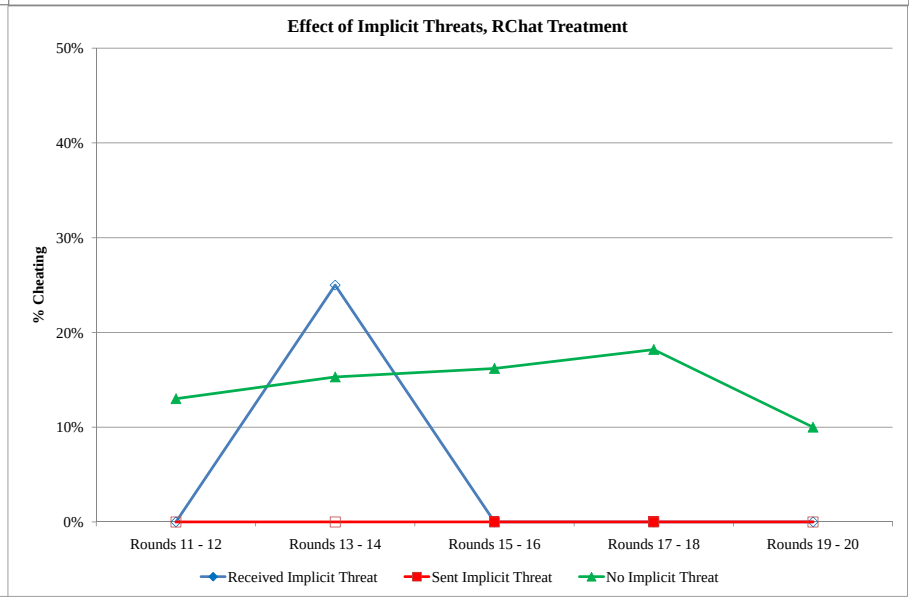


Figure 7: Use and Effect of Categories Related to Other-Regarding Preferences in the PChat Treatment

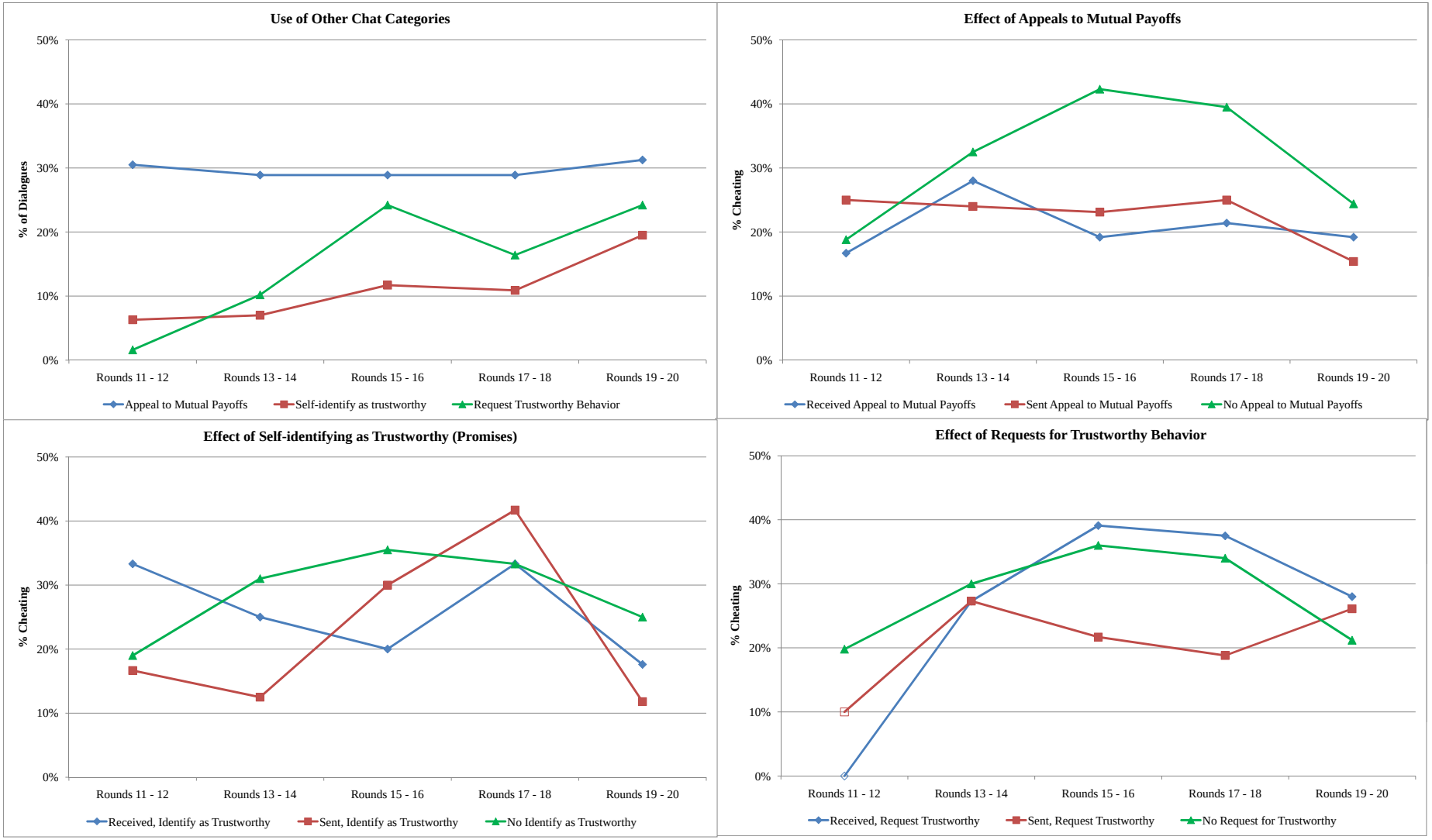


Figure 8: Use and Effect of Categories Related to Other-Regarding Preferences in the RChat Treatment

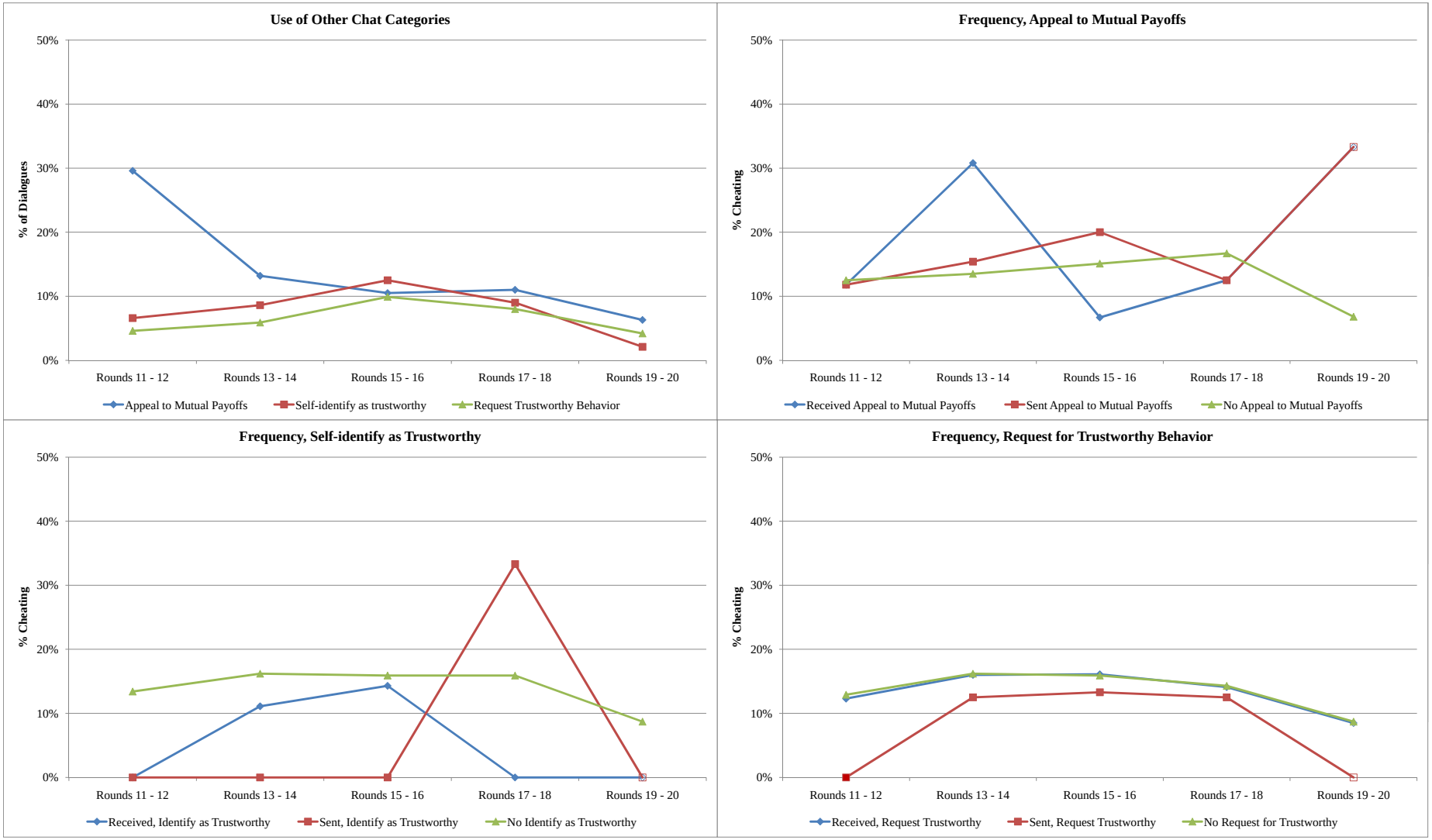


Figure 9: Gains from Cheating and Chat

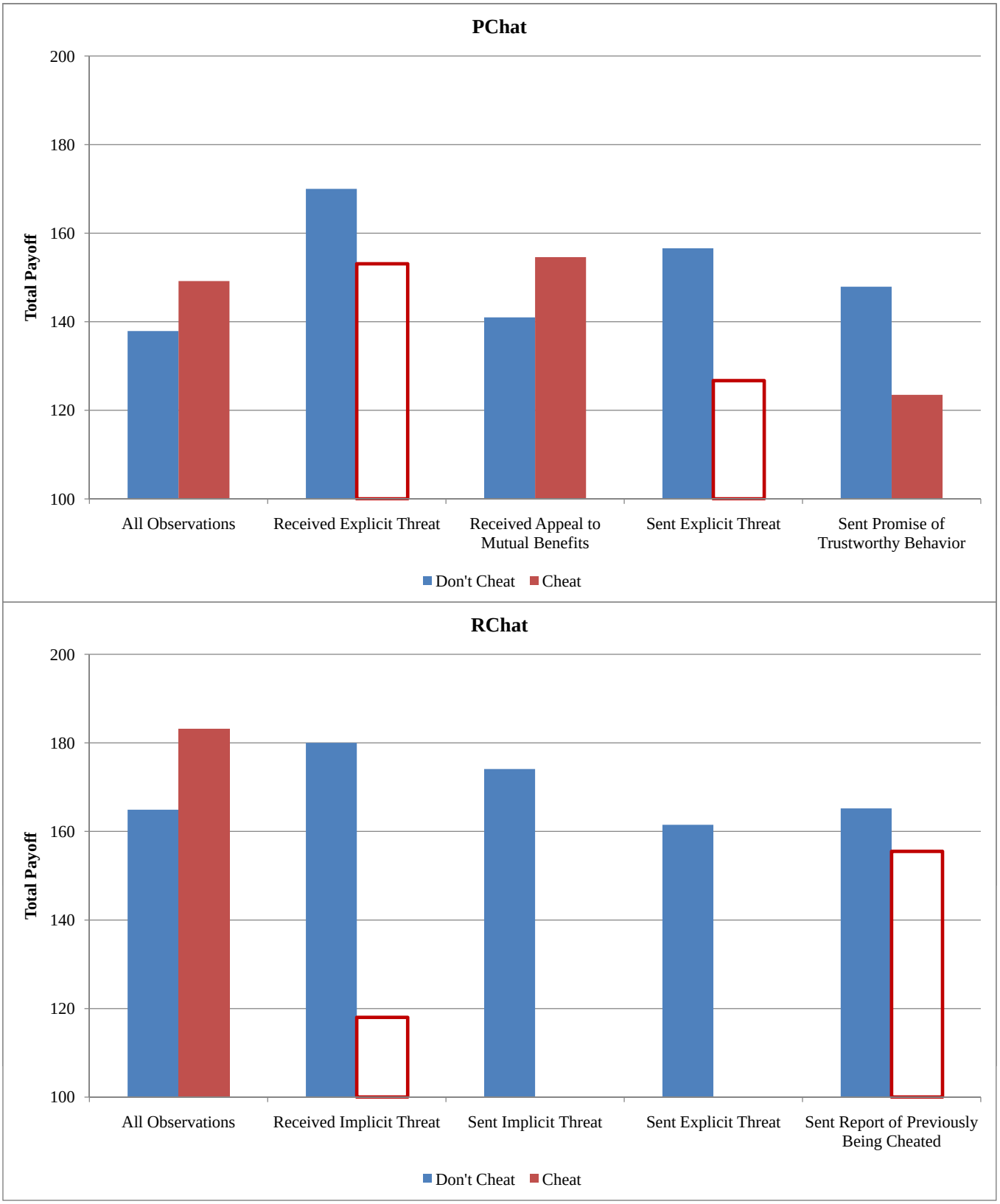


Figure 10: Lagged Effects of Received Messages

