Noname manuscript No.
(will be inserted by the editor)

From Coordination to Double-Crossing: Experiments on the Strategic Behavior of Groups

Marco Castillo · **Greg Leo**

the date of receipt and acceptance should be inserted later

Abstract There is a rich literature on what constitutes an equilibrium that is immune to deviations by coalitions of players. Less is known about the extent to which people actually attempt to coordinate different types of coalitional agreements. We conduct a laboratory experiment designed to test this. Subjects discover and coordinate self-enforcing coalitional deviations including some that are intuitive but more sophisticated than most theories predict. Play in games without communication, where such coordination is impossible, shows that these deviations have a dramatic effect on equilibrium selection. Our experiments highlight the empirical importance of coalition-proof concepts and sophisticated play.

Keywords Group Behavior, Coalitional Incentives, Coordination
JEL Codes C92, C72

1 Introduction

> *"...treachery, though at first sufficiently cautious, yet, in the end, betrays itself."*
>
> - Livy, *History of Rome. Book 44, Chpater 15. (Livy 1850)*

Secret agreements, double-crossing and subterfuge are prevalent in history and culture. Stories of such treachery are often extraordinarily intriguing. These, sometimes complicated, forms of interaction affect our relationships at many levels, from

Marco Castillo
Department of Economics, Texas A&M University and the Melbourne Institute
E-mail: marco.castillo@tamu.edu

Greg Leo
Vanderbilt University Department of Economics, VU Station B #351819 2301 Vanderbilt Place Nashville, TN 37235-1819.
Tel.: 615-322-9697 Fax: 615-343-8495
E-mail: g.leo@vanderbilt.edu

personal to political, as Livy describes in his account of secret negotiations conducted by Rhodes, a Roman ally, with King Perseus who Rome was fighting in the Third Macedonian War (Livy 1850; Dillon and Garland 2013).

These kinds of actions also have a long tradition of theoretical investigation in the game theory literature in the study of strategic coalitional behavior, starting with Aumann (1959). However, there is very little direct empirical evidence on the effect that coalitional deviations have in strategic games[1]. This work is an attempt to close that gap.

We do this by studying behavior in a series of experiments with and without non-binding, preplay communication. The experiments are based on three-player games with multiple Nash equilibria. We use three-player games because they allow us to test the importance of coalitional deviations by exploiting conflicts of interests among subgroups of players.

A natural starting point to analyze multiplayer games with unstructured communication are the various coalition-proof extensions of Nash equilibrium. These theories investigate what equilibria must be considered self-enforcing if coalitions can form[2].

Because these theories leave the details of communication unmodeled, they tend to have sharper predictions than the theories that model communication itself. The downside of not modelling communication explicitly is that some assumptions must be made on what kinds coalitional deviations can undermine agreements.

An agreement between the entire set of players is self-enforing if no coalition has "credible" incentive to deviate through a second agreement. What kinds of further deviations should be considered credible? A common assumption, and that used by Bernheim et al (1987) in the definition of coalition-proof Nash equilibrium [CPNE] is that an agreement by some coalition is credible if no *subcoalitions* have a further credible deviation (individual deviations are always credible).

The rationale for focusing on subcoalitions is that information asymmetries would prevent agreements among coalitions that involve "outsiders"[3]. However, our results suggest that this assumption may not sufficiently capture the sophistication of subjects. Indeed, in our experiment, subjects attempt deviations that involve an outsider.

Moreno and Wooders (1998) appears to be the only other experimental study on coalition-proofness in noncooperative games. They implemented a three-player version of the Matching Pennies game in which two players could gain by correlating their strategies. Because, in their experiment, the coalitional deviation was self-

[1] The theoretical literature provides several examples of the importance of coalitions on strategic behavior. For instance, Konishi et al (1999) discuss common agency games and Delgado and Moreno (2004) discuss equilibrium selection in oligopoly games. Genicot and Ray (2003) show that the extension of risk-sharing and reciprocity can be severely limited by the possibility of coalitional deviations. Dutta and Mutuswami (1997) discuss the efficiency of networks when groups can coordinate their actions and Konishi and Ünver (2006) show relationships between coalition-proof equilibrium and group stability in many-to-many matching problems.

[2] Different concepts differ on what should be considered a credible deviation. See (Bernheim et al 1987; Chakravorti and Kahn 1993; Chakravorti and Sharkey 1993; Greenberg 1990, 1994; Moreno and Wooders 1996).

[3] Bernheim et al (1987) warn of this issue while motivating their coalition-proof Nash equilibrium.

enforcing, the experiment tested players' capacity to implement correlated strategies through cheap talk.

In relation to this, and other experimental studies involving pre-play communication in two-person and three-person games with aligned incentives (see discussion section), our experiment gives further evidence of players' remarkable ability to coordinate their actions in sophisticated ways.

Like Moreno and Wooders (1998), we find that, under communication, correlated strategies better explain our data than independent play. However, our research builds on this work by introducing games with a richer set of coalitional incentives. This allows us to explore in further detail the kinds of deviations that players try to implement.

Our results show that behavior does not always match theoretical predictions. We find that while play converges to the unique strong Nash equilibrium (profiles with no profitable deviation by any coalition) in games without communication, play diverges from it when communication is allowed. This is true even in games where the strong Nash equilibrium coincides with the coalition-proof Nash equilibrium. That is, failure to play the coalition-proof Nash equilibrium is not due to the existence of an equilibrium that is naturally attractive. More importantly, we find that players agree to play an equilibrium that is Pareto-dominated by an egalitarian outcome which can be achieved by a credible deviation under the coalition-proof Nash equilibrium definition.

The evolution of play shows that subjects first learn to play individually self-enforcing agreements and then coalitionally self-enforcing agreements. In the process they learn to implement coalitional deviations that are not predicted by theory: players engage in double-crossing- after two players arrange to deviate from a public agreement, one of those players then secretly arranges a further deviation with the player who was not part of the original deviation. It is the threat of this double-crossing deviation that justifies the stability of the equilibrium in which players arrive, and since double-crossing requires incentive-compatible information transmission, we provide evidence that players are more sophisticated than commonly assumed by theory.

The next section presents the experiments used in this study and discusses theoretical predictions. The section is followed with a description of the experimental procedure. We then report results from the experiment, discuss related literature and conclude.

2 Theory and Hypotheses

This section presents the games used in the experiment and discusses the theoretical predictions of coalition-proof type models. All games are kept as similar as possible in an attempt to minimize idiosyncratic treatment effects. The games are presented in table 1. All games have 3 players: Row, Column and Matrix, and each player has two choices: A or B. Payoffs are arranged in the same order, first number for Row, second number for Column and third number for Matrix.

We use three-player games because this is the minimum environment where coalitional deviations produce results different from coordination on self-enforcing Pareto efficient outcomes.

Game 1

	A			B	
	A	B		A	B
A	$\frac{3}{2},\frac{3}{2},\frac{3}{2}$	0,0,0	A	0,0,0	4,2,1
B	0,0,0	0,0,0	B	0,4,0	2,2,2

Game 2

	A			B	
	A	B		A	B
A	$\frac{3}{2},\frac{3}{2},\frac{3}{2}$	0,0,0	A	0,0,0	4,1,2
B	0,0,0	2,2,2	B	0,0,0	0,0,4

Game 3

	A			B	
	A	B		A	B
A	$\frac{3}{2},\frac{3}{2},\frac{3}{2}$	0,0,0	A	0,0,0	4,1,3
B	0,0,0	2,2,2	B	0,0,0	0,0,4

Table 1: Experimental Treatments

The three games are slight variations of each other. For instance, all the games share the same Nash equilibria in pure strategies: (A, A, A), (B, A, B) and (A, B, B) and a single mixed-strategy equilibrium in which Row plays A while Column and Matrix mix[4].

Among the equilibria, (B,A,B) is Pareto-dominated by all other equilibria in Games 2 and 3. In game 1, B is a weakly dominated for Row. The remainder of the equilibria in each game are those of the assymetric coordination game faced by Column and Matrix conditional on Row playing A.

Experimental research on the effects of nonbinding preplay communication in coordination games with asymmetric payoffs shows that people strive to break the asymmetry created by opposite incentives (Camerer 2003, section 7.3). One-sided communication tends to ameliorate the coordination problem while two-sided communication leaves it unchanged. There is no *a priori* reason to believe that pre-play communication will favor a particular equilibria over the other in our experiments.

4 The mixed equilibria involve Column playing A with probability 2/5 for Game 1, 4/7 for Game 2, and 2/3 for Game 3. For Matrix the probabilities are 4/7 for Game 1, 2/5 for Game 2, and 2/5 for Game 3. The payoffs are such that if players were to play a mixed-strategy equilibrium, the predicted frequency of play of equilibrium (A, A, A) is between 23 percent and 27 percent and the predicted frequency of play of (A, B, B) is between 20 percent and 26 percent.

However, we note that (A, A, A) produces symmetric payoffs while neither (A, B, B) nor the mixed-strategy equilibria do (see below). We will refer to (A, A, A) as the symmetric Nash equilibrium[5].

A first hypothesis is that communication will produce more equilibrium play. In a sense, this requirement is minimal since players can use more complicated strategies once communication is allowed. In games with non-binding pre-play communication players can coordinate or even correlate their strategies[6].

A second hypothesis is that players will coordinate on equilibria where no coalition has a profitable deviation. For instance, while (B, A, B) is a Nash equilibrium of all the games, it can be deviated from by coalitions of players. In Game 1, Row and Matrix can switch to A instead and secure $\frac{3}{2}$ for all players. In Game 2 and Game 3, this equilibrium can also be challenged by the coalition of all players.

The mixed-strategy equilibria are also not robust to deviations of coalitions. Indeed, in all games, the mixed strategy equilibria give strictly less to Column and Matrix than either the symmetric Nash equilibrium or (A, B, B).

Finally, note that the symmetric Nash equilibrium is not robust to deviations by coalitions either. In game 1, all players can agree to play (B, B, B) instead, and in Game 2 and Game 3, they can agree to play (B, B, A). These deviations strictly improve the payoffs of all players. We will refer to this alternative (B,B,A) in games 2 and 3 as the symmetric Pareto efficient outcome.

This leaves (A,B,B) as the sole equilibrium in which no coalition can jointly improve their payoffs. The process of elimination of equilibria just described defines strong Nash equilibrium (SNE) of Aumann (1959). SNE is similar to the core concept of cooperative game theory (Greenberg 1994) and fails to exist for many games of economic interest (e.g., games with a Prisoner's Dilemma structure). We will refer to (A,B,B) as the strong Nash equilibrium in all games.

The main criticism of SNE is that it considers deviations that are not immune to further deviations. Take the case of a deviation from the symmetric Nash equilibrium to the symmetric Pareto efficient outcome in Game 1. After the deviation, Row or Column can deviate to (B, A, B) or (A, B, B) making player Matrix worse-off. Therefore, the original deviation from the symmetric Nash equilibrium is not credible in this game. Matrix would be unwise to follow it.

This criticism is summarized in the coalition-proof Nash equilibrium concept (Bernheim et al 1987). An agreement is coalition-proof Nash equilibrium (CPNE) if it is immune to *credible* deviations by coalitions. The concept is nested. Taking the actions of others as given, a player's best response is always credible. Two players' deviations are credible if they are Nash equilibria of the game resulting from keeping the choices of the non-deviating players fixed. So, in three-player games, an agree-

[5] Even though technically (B,A,B) is a Nash equilibrium offering symmetric payoffs in games 2 and 3.

[6] For instance, players can agree to play in the following manner: each player announces either A or B with equal probability, whatever the announcement, Row always plays A. If Column's and Matrix's announcements coincide they both play A, and if their announcement disagree they both play B. The expected payoff of this agreement, in Game 1, is 2.75 for player Row, 1.75 for player Column and 1.25 for player Matrix. These payoffs are higher than the payoffs resulting from the equilibrium in mixed strategies. Moreover, this agreement is implementable as a Nash equilibrium of the game with pre-play communication. Other agreements would be possible if mediated communication is possible (Myerson 1997).

ment is CPNE if it is immune to credible deviations and there is no other immune agreement that makes all three players better off.

In Game 1, there are two CPNE: the symmetric Nash equilibrium and the strong Nash equilibrium.

In Game 2 and Game 3, the unique CPNE is the strong Nash equilibrium [7].

The symmetric Nash equilibrium is not CPNE in games 2 and 3. A deviation by Row and Column to the symmetric Pareto efficient outcome is credible. If Matrix does not know that the deviation has taken place he cannot deviate to (B, B, B). More importantly, neither Row nor Column have an incentive to further deviate from the symmetric Pareto efficient outcome once they reach it[8].

A third hypothesis is then that subjects will play according to CPNE: the symmetric Nash equilibrium should not be played in Game 2 and Game 3.

The nestedness of coalitional deviations in the CPNE concept insures that all deviators share knowledge of the deviations that have already taken place. This implicitly rules out the possibility that players might transmit truthful information about other players' deviations. Thus, CPNE rules out what we call *double-crossing*.

The recommendation of CPNE for Game 2 and Game 3 depends critically on the assumption that only the members of a deviating coalition can deviate further. However, Row does have an incentive to deviate further with Matrix's help. That is, Row and Matrix have an incentive to double-cross Column.

Why would Row try to deviate from the symmetric Pareto efficient outcome? And, why would Matrix cooperate if he cannot verify that a previous deviation has occurred? Row would like to reveal that the deviation has occurred because, once Column has committed to play B, the remaining game between Row and Matrix has a unique Nash equilibrium, the strong Nash equilibrium. And this equilibrium gives Row his highest possible payoff in Game 2 and Game 3.

Row does not have an incentive to lie about the occurrence of this deviation. Suppose, on the contrary, that Row convinces Matrix that Column is playing B when he is not. If Matrix believes this lie and acts accordingly by playing dominant strategy B, everyone will earn zero regardless of Row's action. In this way, Matrix understands Row has no incentive to lie about Column playing B and should take the message seriously.

Given the possibility of double-crossing, Column should consider a deviation from the symmetric Nash equilibrium to the symmetric Pareto efficient outcome to be unacceptable[9].

[7] All strong Nash equilibria are CPNE. They have no credible deviations since they have no profitable deviations.

[8] The concept of coalition proof equilibrium [CPE] extends CPNE to the case in which correlated strategies are available (Moreno and Wooders 1996). CPE also predicts that only (A,B,B) will survive in Game 2 and Game 3.

[9] To our knowledge, the only concepts that extend CPNE to the case in which these deviations are possible are Chakravorti and Sharkey (1993) consistent coalition proof equilibrium (CCPE) and Chakravorti and Kahn (1993) universal coalition proof equilibrium (UCPE). Only UCPE captures the logic of our argument.

Universal coalition proof suggests that after a deviation has occurred, some of the deviators could invite some new players to deviate further. The new players would play along provided that any move that could harm them would harm at least one of the original deviators. This condition would give new players peace

We hypothesize that double-crossing takes place making (A, A, A) robust to deviations by coalitions.

In summary, strong Nash equilibrium predicts that in all games only strategy profile (A, B, B) will be selected. Coalition-proof equilibrium also predicts that strategy profile (A, B, B) will be selected in Game 2 and Game 3, but does not rule out (A,A,A) in Game 1. However, profile (A, A, A) is also not ruled out in either Game 2 or Game 3 if subjects engage in double-crossing.

3 Experimental Design

We implemented experimental sessions of the three games described in the previous section with and without communication at the University of Wisconsin-Madison and at the Georgia Institute of Technology. A total of 90 students participated in four sessions consisting of 21 or 24 participants in Wisconsin-Madison (Game 1 and Game 2 with communication). Except for session 1 of Game 1, which consisted of 10 rounds, all the sessions consisted of 13 rounds of play. All the participants were told that they were going to play 10 to 15 rounds of the game as time permitted[10]. A total of 180 students participated in 10 sessions consisting of 18 participants at Georgia Tech. All sessions consisted of 15 rounds of play and all participants were told that they were going to play 10 to 20 round of the game as time permitted[11]. In total there are 270 subjects. Each experimental session lasted about an hour and a half, and each subject received an average of \$20.35 (std = \$13.29), plus a \$6 show-up fee.

Each round of play in the games with communication consisted of three stages, a communication stage that lasted at most three minutes, a decision-making stage and a payoff-action-review stage without time constraints[12]. At the beginning of the experiment, each subject was randomly assigned a color (Red, Blue or Green), corresponding to the role of Row, Column, and Matrix. Each player kept the same color/role for the entire experiment. In each round of play, groups consisting of one player of each color were randomly formed with the exception that no exact group of three players was ever in sequence. The sessions without communication used the same procedure with the communication stage removed.

In sessions with communication, we read the experimental instructions aloud while subjects followed along on printed copies. After this, they faced the screen shown in appendix figure 3. Here, they were informed which color/role they had been assigned. The panels in the screen changed to signal the change of the stage of

of mind that they are not been deceived. However, (A, A, A) is not UCPE for Game 2. The reason being that at (B, B, A) there is not a deviation that makes both player Row and player Column strictly better off. (A, A, A) is a UCPE in Game 3.

[10] Instructions can be found here:
`https://my.vanderbilt.edu/gleo/files/2015/11/game3comm.pdf`
`https://my.vanderbilt.edu/gleo/files/2015/11/game3no.pdf`.

[11] The choice of the last round was made to prevent last round effects. A random stopping rule could have been used, but our method insures comparable session sizes. We did not find any evidence that subjects anticipated the end of a session.

[12] The sessions in Madison had a time constraint for the review of payoff stage that lasted 20 seconds. The constraint was never binding.

the game. For instance, the upper left window changed to show which round of play was being played. The payoff matrix and subject's role were always visible.

Payoffs were presented in the corresponding color to make the reading of the matrix easier. Subjects' colors were determined when the game started. Indeed, the experimenter did not know ahead of time which role each person was going to be assigned.

The right-side panel changed throughout the stages of the experiment. Figure 3 shows the window as it appeared during communication-stage. This window changed after three minutes to a menu in which subjects were asked to make their decisions. After all subjects had submitted their decisions, the upper right window changed again to show the outcome of that round, including the decisions and payoffs for each player.

After all players were done reviewing the results, subjects were regrouped and a new communication stage started for a new round of play. The-right bottom panel in figure 3 shows the messages sent by other players. Messages could either be sent publicly to the other two players or be sent privately to only one of them.

In the no-communication sessions, all the procedures were the same except that subjects did not have three minutes to send messages. Also, in figure 1, the right-hand panel was only a decision-making screen[13].

4 Results

4.1 Basic Results

This section presents a general overview of play in the games with and without communication. Table 2 presents the frequency with which each strategy profile was chosen. Each of these follows the presentation of the game matrix in table 1 to facilitate the presentation. The left column of table 2 are the games with communication and the right are the games without.

			Blue Action / Green Action							
			A/A	A/B	B/A	B/B	A/A	A/B	B/A	B/B
			No Communication				Communication			
Red Action	A	Game 1	0.022	0.111	0.156	0.400	0.427	0.088	0.111	0.146
	B		0.022	0.039	0.078	0.172	0.018	0.023	0.088	0.099
	A	Game 2	0.033	0.094	0.172	0.628	0.235	0.144	0.075	0.123
	B		0.000	0.011	0.006	0.056	0.019	0.264	0.008	0.133
	A	Game 3	0.017	0.128	0.161	0.639	0.306	0.022	0.078	0.328
	B		0.000	0.006	0.017	0.033	0.011	0.133	0.011	0.111

Table 2: Distribution of Play

The first observation is that, in the absence of communication, subjects tend to play the strong Nash equilibrium of the games. This fact is clearer in Game 2 and Game 3.

[13] All the experiments used z-tree (Fischbacher 2007).

Statistical tests show that the play in Game 2 and Game 3 without communication is not different, but that play in Game 1 is different to Game 2 and Game 3 combined[14]. This seems to be due to attempts to coordinate on the play of (B,B,B) in Game 1[15]. This suggests that strategy profile (B,B,B) is more focal in Game 1 than strategy profile (B,B,A) is in games 2 and 3. Indeed, conditional on Row playing A, the distributions of play by Column and Matrix are not statistically different from each other[16].

Communication has a significant and striking effect on the distribution of play in each game[17]. This result is consistent with previous experimental evidence[18].

The concept of correlated equilibrium (Aumann 1974; Myerson 1982) is appropriate whenever players can communicate prior to play. For instance, players can alternate between multiple Nash equilibria using unmediated communication[19]. We should then expect that correlated equilibria better describe the average play in the games with communication than in games without communication.

Appendix table 5 presents maximum likelihood estimates of two models of aggregate play. The first column of each game fits an unconstrained multinomial model of the decisions in each game. The model does not restrict behavior to be optimal. The second column of each game presents maximum likelihood estimates of a model that allows for correlated play. Importantly, these estimates take into consideration the theoretical restrictions of correlated Nash equilibrium[20].

The bottom panel of appendix table 5 reports the likelihood ratio test of the hypothesis that the correlated equilibrium inequalities are valid restrictions both for the entire sample and for the second-half of rounds (8 on). These restrictions are never rejected in games with communication past round 8, but are rejected in games 1 and 3 without communication. Since correlated equilibrium is a generalization of Nash equilibrium, a rejection of the correlated equilibrium model also implies rejection of Nash equilibrium. Thus, these results suggest that communication had an role in helping subjects coordinate equilibrium play.

Two final patterns of behavior are worth remarking: the frequency of play of strategy profile (A,B,B) in games without communication (40, 63, and 64 percent in games 1,2,3 respectively), and the frequency of play of strategy profile (A,A,A) in games with communication (43, 24, and 30 percent in games 1,2,3 respectively).

[14] The test of equality of distribution of play is $\chi^2(7) = 38.68$, p-value = 0.000 between Game 1 and Game 2, $\chi^2(7) = 42.77$, p-value = 0.000 between Game 1 and Game 3 and $\chi^2(6) = 4.32$, p-value = 0.634 between Game 2 and Game 3.

[15] The probability that subjects coordinate on the highest payoff strategy profile is 17 percent in Game 1 and 1 percent in Game 2 and Game 3. The test that this probability is equal in Game 1 and games 2 and 3 combined is t-test = -7.7817, p-value = 0.000.

[16] $\chi^2(6) = 5.55$, p-value = 0.475.

[17] The test of equality of distribution is 91.95 (p-value = 0) for Game 1, 203.65 (p-value = 0.000) for Game 2 and 114.15 (p-value = 0.000) for Game 3.

[18] See Kagel and Roth (1995, chapter 3.2), Camerer (2003, chapter 7.2), Crawford (1998), Blume et al (2001), Clark et al (2001), Charness (2000), Blume and Ortmann (2007) among many.

[19] Gerardi (2004) discusses sufficient conditions for groups of players to implement correlated equilibria in the absence of a mediator.

[20] We relax the correlated equilibrium to allow for mistakes. In particular, we consider that a restriction is violated if it exceeds 0.05 in expectation. This ε-correlated equilibrium allows us to deal with mistakes and those cases where a player has a weakly dominant strategy as player Row in Game 1.

The first pattern of behavior is remarkable because players coordinate on an asymmetric Nash equilibrium without the aid of communication. The second pattern is remarkable because strategy profile (A,A,A) is dominated in all three games by another strategy profile. Thus, the pattern is both non-salient without communication and cannot be explained by fairness alone. In the next section, we bring together the strategies with preplay communication to explain these patterns. In particular, we present evidence showing that subjects attempt to implement coalitional deviations.

4.2 Examples of Deviations and Double-Crossing

Before looking at the joint distribution of messages and play, we present two examples of the dialogues that took place during the games. Both examples demonstrate that subjects learned to use the incentives of the game to their advantage. The examples belong to Game 3, but similar examples are also found in Game 2.

Table 3 shows an example of a successful coordination by Row and Column on the play of the Pareto efficient outcome. As the dialogue shows, Row and Column make an effort to convince Matrix that they will play A. Importantly, Matrix recognizes that Row and Column have an incentive to coordinate on the symmetric Pareto efficient outcome once the symmetric Nash equilibrium has been agreed. He also recognizes that enough confidence on a deviation to the symmetric Pareto efficient outcome will prompt further deviations.

Round 9, Game 3, played (B, B, A), payoff (200, 200, 200)
Row (To All): what should we do? *Column (To All)*: a a a *Row (To All)*: sounds good to me *Column (To All)*: so Matrix doesn’t screw us over *Matrix (To All)*: alrigh, straight up, im picking A, because otherwise somebody goes crappy and i end up with 0 somehow *Row (To All)*: I hear that I have already been screwed over by Matrix *Row (To Column)*: hey why don’t we pick b so that we all end up making 200each... as well as the Matrix player *Column (To Row)*: are you sure you trust Matrix? *Row (To Column)*: what do you want to do *Column (To Row)*: I’ve been screwed over that way before... I think we should stick to a *Row (To Column)*: I mean if he thinks that we are picking a he doesn’t get paid unless he pick a...see so don’t let him know that he COULD screw us over *Row (To All)*: im going with a as well *Row (To Column)*: im really picking B *Matrix (To All)*: have you guys decided? don’t tell me if its 200, so then I can’t screw you over

Table 3: Dialog Example 1: Coordination on B,B,A.

Table 4 presents an example of double-crossing. Row and Column deviate from the symmetric Nash equilibrium to the egalitarian Pareto efficient outcome and tell

Matrix that they both will play A. Row informs Matrix that this deviation has taken place and that both can further deviate from the egalitarian Pareto-efficient outcome to the strong Nash equilibrium. Row succeeds in implementing the unique coalition-proof Nash equilibrium by means of this deception.

Round 14, Game 3, play (A, B, B), payoffs (400, 100, 300)

Row (To All):150s?
Column (To All): sure
Column (To Row): Row, if we make green think we're going for 150, we can both choose B and get 200
Row (To Column): I was going to tell u that too
Row (To All): 150s
Matrix (To All): Fine, 150s
Row (To Column): so we are going for 200s? Rite
Row (To Matrix): hey Matrix
Matrix (To Row): Yeah, whats up
Column (To Row): make Matrix think it's 150 though
Row (To Matrix): column is actually going for 200 200 200... so play 413...but don't tell him
Matrix (To Row): Gotcha
Column (To All): 150 ensures no cheating
Row (To Column): 222 nice...

Table 4: Dialog Example 2: Double-Crossing.

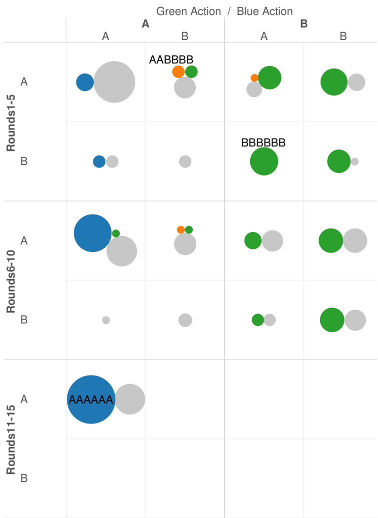
4.3 Results: Deviations and Double-Crossing

Figure 2 below provides an overview of the evolution of agreements and play in our games. This figure provides the relative frequency of key message/play combinations for each game and 5-round interval. Each "bubble" represents a particular message. The coloring is consistent over the games. For instance, the royal blue bubble always represents the messages "AA,AA,AA". The cell where the bubbles appears gives the action profile that resulted. For instance, the royal blue bubbles in the cell A,A,A is represents the combination of the "AA,AA,AA" massages with play A,A,A. That is, this bubbles represents coordinating successfully on A. The size of each bubble represents the relative frequency of that message/play combination in the 5-round interval for that game.

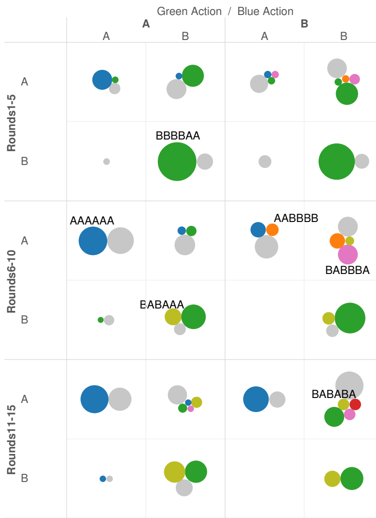
The profiles of communications presented in this figure are:

– AA,AA,AA- coordination on the symmetric Nash Equilibrium.
– AA,BB,BB- coordination on the strong Nash equilibrium.
– BB,BB,AA- coordination on the symmetric Pareto efficient outcome.
– BA,BA,AA- Row and Column deviate from the symmetric Nash to the symmetric Pareto outcome.
– BA,BA,BA- Double crossing profile. From the symmetric Nash equilibrium, Row deviates with Column to the Pareto outcome and then enlists Matrix to further deviate to the strong Nash equilibrium.

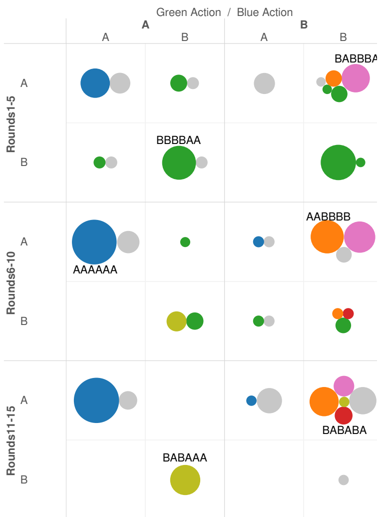
– BA,BB,BA- Row and Matrix deviate from the symmetric Pareto outcome to the strong Nash equilibrium.



Game 1



Game 2



Game 3

Fig. 2: Joint Distribution of Actions and Messages

There are some general patterns in the data. First, in all games, early attempts to agree on the symmetric Pareto efficient outcome often ended in failure either due to deviations by a single individual, or in the case of game 3, by the coalition of Row and Matrix to the strong Nash equilibrium. In games 1 and 3, attempts to coordinate on the symmetric Pareto efficient outcome almost disappear by the end. In game 2 they remain, although much of this can be attributed to a single exceedingly cooperative session (out of four). On the other hand, agreements about either of the Nash equilibria were almost always successful.

After several rounds in games 2 and 3, Row and Column begin attempting deviations from the symmetric Nash equilibrium to the symmetric Pareto efficient outcome. Yet, by the end of all games, a large proportion of groups play the symmetric Nash equilibrium by explicit agreement. What is supporting these agreements to play AAA in games 2 and 3? Or equivalently, why are there not more attempts to deviate to the symmetric Pareto efficient outcome as would be predicted by the model of Coalition-proof Nash equilibrium?

Technically, repeated-game effects could play a role. As seen in the second dialog example above, Matrix players were likely well aware of the incentives for Row and Column to deviate. With enough confidence that the deviation would take place, Matrix players would have incentive to gamble on playing B.

In game 2 many attempts by Row and Column to deviate end in this way- with Matrix deviating to B in anticipation of Row and Column's agreement[21]. However, this does not appear to have been a factor at all in game 3. Attempts to deviate were almost always successful.

We argue, instead, that the possibility of double-crossing at least contributes to the stability of the agreement on the symmetric Nash equilibrium. In fact, the explicit double-crossing profile "AB,AB,AB" with play ABB shows up towards the middle of Game 3 and is more apparent towards the end. Of course, the effect of the *possibility* double-crossing does not require it to happen at all. However, the fact that it does happen lends strong evidence to our hypothesis.

To make these patterns more clear, we estimate a model of noisy messages and play around the key agreements and deviations[22]. The estimates are reported in appendix table 6 and the resulting predictions about the probabilities of each type of agreement or deviation (both at round 5, 10 and 15) are provided in appendix table 7.

The key profiles we estimate are the same as presented in 2 and listed above. This table confirms many of the patterns on the evolution of play apparent in figure 2. Agreements on the symmetric Pareto efficient outcome are prevalent in early play of all three games. Explicit agreements and associated deviations from this profile nearly disappear towards the end of the rounds in all games.

At round 15, almost all play in game 1 is predicted to be the symmetric Nash equilibrium. In game 2, play is predominately a mixture of the symmetric Nash and devi-

[21] It appears possible that Row often even anticipated Matrix anticipating the deviation, leading to the strong Nash outcome- we might call this tacit double-crossing.

[22] In particular, we consider that players try to implement a strategy profile (x,y,z) by sending message (m_1,m_2,m_3) but they by might make mistakes by either sending message $(\hat{m}_1,\hat{m}_2,\hat{m}_3)$ and/or playing strategy profile $(\hat{x},\hat{y},\hat{z})$. Each of these errors (message or play by each player) is made independently with probability ε. If players implement strategy profile x_1 with probability α and strategy profile x_2 with probability $1-\alpha$ the likelihood of observing a pattern of communication and play ω is $Pr(\omega|x_1,\alpha,\varepsilon)\alpha + Pr(\omega|x_2,\alpha,\varepsilon)(1-\alpha)$. The appropriately modified likelihood expression is what we maximize to produce the results in appendix table 6

ations to the symmetric Pareto outcome. In game 3, there is a mixture of agreements of strong Nash, and symmetric Nash agreements. While the proportion of agreements on the symmetric Nash is not significantly different than in game 2, there are significantly fewer[23] deviations to the the symmetric Pareto outcome. As we have argued above, this may be due to the threat of double-crossing which appears as a significant profile for the first time at the end of this game.

5 Discussion

5.1 Related Literature

The experimental literature on the role of communication in two-player games is vast. In coordination games, previous experimental research shows that players try, and many times succeed, in coordinating on efficient equilibria (see: Crawford (1998); Charness (2000); Burton and Sefton (2004)). For instance, Blume et al (2001) show that coordination is possible even when messages do not have an *a priori* meaning. However, communication may fail to promote coordination in games where preferences between players are not aligned, as in the Battle of Sexes game (Cooper et al 1989). In other situations, risks associated with coordination failure might prevent successful coordination (Clark et al 2001; Burton and Sefton 2004).

Games with preplay communication and three or more players have focused primarily on coordination in games with Pareto-ranked equilibrium. For instance, Blume and Ortmann (2007) show preplay communication increases efficiency in median and minimum games. Weber (2006) shows that groups can solve coordination problems by growing slowly. In a meta-study of step-level public goods games (each with more than three players), Croson and Marks (2000) find that communication promotes efficiency. However, the richness of coalition proof concepts is not fully realized in such games. The incentives of the entire group of players, each sub-group and each player are fully aligned.

5.2 Conclusions

This research set out to investigate the relevance of coalitional deviation as a selection criteria in strategic games. Previous experiments have shown that communication in two-player games increases coordination and could help avoid Pareto inefficient equilibria (Charness 2000; Burton and Sefton 2004). Talk might be cheap, but it is not worthless. We confirm that nonbinding preplay communication affects the play of games. Subjects are able to navigate through incentives and find agreements that are self-enforcing. Moreover, we find that coalitional deviations are important in determining what agreements are self-enforcing.

Paradoxically, we find that without communication, play converges strongly to a coalition-proof (and strong) Nash equilibrium, but that with communication play moves away from it. Our results show that this cannot be completely explained by

[23] Against a one-sided alternative.

fairness considerations. Instead, we find that coordination on complicated coalitional deviations might play a role. Indeed, an important finding of our experiments is that once subjects understand the incentives of the games at hand, they can engage in sophisticated play. Either players openly plan on double-crossing each other or engage in incentive-compatible information transmission. We find that coalition-proofness might fail not only because of bounded rationality, but because equilibrium models assume too little rationality. The kind of deceptions observed in our games are consistent with the arguments in Crawford (1998) that boundedly rational agents might lie and get away with it sometimes. Interestingly, we find an evolution towards more and more involved forms of deception.

Our results show that the answer to the question of what constitutes a coalition-proof agreement is a complicated one. As previous experimental research has shown, subjects demonstrate an enormous ability to coordinate their actions. Our experiments encourage further study to better understand these issues. In particular, how self-signaling messages may permit more sophisticated forms of coordination than are commonly assumed in theoretic work.

References

Aumann R (1959) Acceptable Points in General Cooperative n-Person Games. Contributions to the theory of games 4:287–324

Aumann RJ (1974) Subjectivity and correlation in randomized strategies. Journal of mathematical Economics 1(1):67–96

Bernheim B, Peleg B, Whinston MD (1987) Coalition-Proof Nash Equilibria I. Concepts. Journal of Economic Theory 42:1–12, DOI 10.1016/0022-0531(87)90099-8

Blume A, Ortmann A (2007) The effects of costless pre-play communication: Experimental evidence from games with Pareto-ranked equilibria. Journal of Economic Theory 132(1):274–290, DOI 10.1016/j.jet.2005.03.001

Blume A, DeJong DV, Kim YG, Sprinkle GB (2001) Evolution of Communication with Partial Common Interest. Games and Economic Behavior 37(1):S. 79–120, DOI 10.1006/game.2000.0830

Burton A, Sefton M (2004) Risk, pre-play communication and equilibrium. Games and Economic Behavior 46(1):23–40, DOI 10.1016/S0899-8256(03)00115-5

Camerer C (2003) Behavioral Game Theory: Experiments in Strategic Interaction. Insights in decision making: A tribute to Hillel J Einhorn p 544, DOI 10.1024/0301-1526.32.1.54

Chakravorti B, Kahn C (1993) Universal coalition-proof equilibrium: Concepts and applications. Papers

Chakravorti B, Sharkey B (1993) Consistency, un-common knowledge and coalition-proofness

Charness G (2000) Self-Serving Cheap Talk: A Test Of Aumann's Conjecture. Games and Economic Behavior 33(2):177–194, DOI 10.1006/game.1999.0776

Clark K, Kay S, Sefton M (2001) When are Nash equilibria self-enforcing? An experimental analysis. International Journal of Game Theory 29:495–515, DOI 10.1007/s001820000054

Cooper R, Dejong DV, Forsythe R, Ross TW (1989) Communication in the battle of the sexes game: some experimental results. RAND Journal of Economics 20(4):568–587, DOI 10.2307/2555734

Crawford V (1998) A Survey of Experiments on Communication via Cheap Talk. Journal of Economic Theory 78(2):286–298, DOI 10.1006/jeth.1997.2359

Croson R, Marks MB (2000) Step returns in threshold public goods: A meta-and experimental analysis. Experimental Economics 2(3):239–259

Delgado J, Moreno D (2004) Coalition-proof supply function equilibria in oligopoly. Journal of Economic Theory 114(2):231–254, DOI 10.1016/S0022-0531(03)00123-6

Dillon M, Garland L (2013) Ancient Rome: A Sourcebook. Routledge

Dutta B, Mutuswami S (1997) Stable Networks. Journal of Economic Theory 76(2):322–344, DOI 10.1006/jeth.1997.2306

Fischbacher U (2007) z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10(2):171–178

Genicot G, Ray D (2003) Group formation in risk-sharing arrangements. Review of Economic Studies 70(1):87–113, DOI 10.1111/1467-937x.00238

Gerardi D (2004) Unmediated communication in games with complete and incomplete information. Journal of Economic Theory 114(1):104–131

Greenberg J (1990) The Theory of Social Situations: An Alternative Game-Theoretic Approach. Cambridge University Press

Greenberg J (1994) Coalition structures. Handbook of Game Theory with Economic Applications 2:1305–1337, DOI 10.1016/S1574-0005(05)80069-4

Kagel JH, Roth AE (1995) The Handbook of Experimental Economics, vol 78. DOI 10.2307/1243316, 0701184v1

Konishi H, Ünver MU (2006) Credible group stability in many-to-many matching problems. Journal of Economic Theory 129(1):57–80, DOI 10.1016/j.jet.2005.02.001

Konishi H, Le Breton M, Weber S (1999) On Coalition-Proof Nash Equilibria in Common Agency Games. Journal of Economic Theory 85(1):122–139, DOI 10.1006/jeth.1998.2481

Livy (1850) History of Rome by Titus Livius, books thirty-seven to the end, with the epitomes and fragments of the lost books. Henry G. Bohn

Moreno D, Wooders J (1996) Coalition-Proof Equilibrium. Games and Economic Behavior 17(1):80–112, DOI 10.1006/game.1996.0095

Moreno D, Wooders J (1998) An Experimental Study of Communication and Coordination in Noncooperative Games. Games and Economic Behavior 24(1–2):47–76, DOI 10.1006/game.1997.0624, URL http://www.sciencedirect.com/science/article/pii/S0899825697906242

Myerson RB (1982) Optimal coordination mechanisms in generalized principal–agent problems. Journal of mathematical economics 10(1):67–81

Myerson RB (1997) Game Theory. Harvard University Press, URL `https://books.google.com/books?id=E8WQFRCsNr0C{\&}pgis=1`

Weber Ra (2006) Managing growth to achieve efficient coordination in large groups. American Economic Review 96(1):114–126

6 Appendix

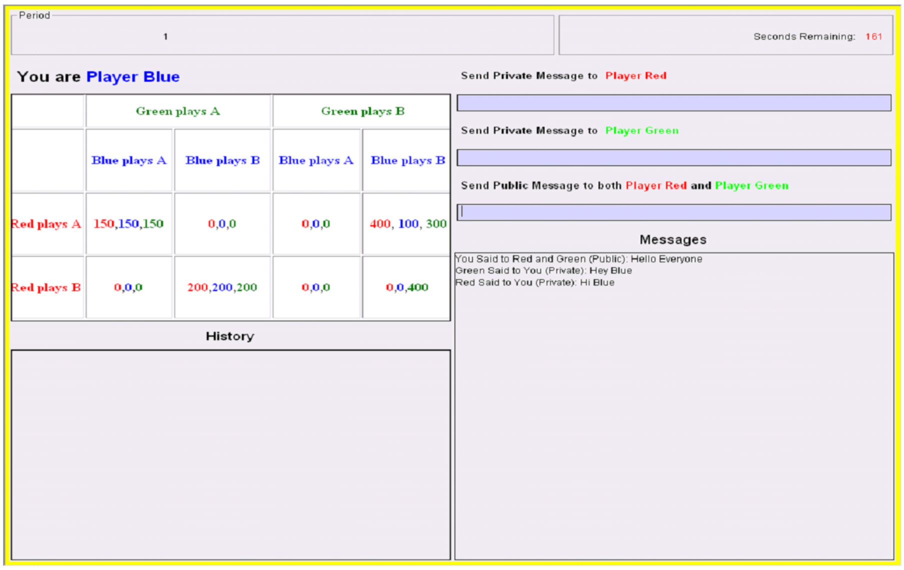


Fig. 3: Experiment Screen for the Communication Stage

Table 5: Maximum Likelihood Estimates, Unconstrained and Correlated Equilibrium

	No communication						Communication					
Strategy	Game 1		Game 2		Game 3		Game 1		Game 2		Game 3	
Profile	Uncons.	e-CE	Uncons.	e-CE	Uncons.	e-CE	Uncons.	e-CE	Uncons.	e-CE	Uncons.	e-CE
$A_1A_2A_3$	0.022	0.070	0.033	0.076	0.017	0.078	0.427	0.471	0.235	0.367	0.306	0.307
$A_1A_2B_3$	0.156	0.180	0.172	0.163	0.161	0.167	0.111	0.123	0.075	0.076	0.078	0.078
$A_1B_2A_3$	0.111	0.100	0.094	0.064	0.128	0.053	0.088	0.097	0.144	0.100	0.022	0.022
$A_1B_2B_3$	0.400	0.494	0.628	0.653	0.639	0.666	0.146	0.161	0.123	0.125	0.328	0.329
$B_1A_2A_3$	0.022	0.004	0.000	0.000	0.000	0.000	0.018	0.005	0.019	0.017	0.011	0.010
$B_1A_2B_3$	0.078	0.103	0.006	0.006	0.017	0.017	0.088	0.097	0.008	0.008	0.011	0.011
$B_1B_2A_3$	0.039	0.028	0.011	0.017	0.006	0.004	0.023	0.026	0.264	0.201	0.133	0.156
$B_1B_2B_3$	0.172	0.022	0.056	0.021	0.033	0.015	0.099	0.021	0.133	0.107	0.111	0.087
log Likelihood	-305.5	-352.9	-210.8	-218.4	-202.0	-216.9	-291.4	-306.6	-676.3	-694.5	-292.3	-293.2
Observations	180	180	180	180	180	180	171	171	375	375	180	180
	$\chi^2(6)$	p-value	$\chi^2(6)$	p-value	$\chi^2(6)$	p-value	$\chi^2(6)$	p-value	$\chi^2(6)$	p-value	$\chi^2(6)$	p-value
L.R. test	94.9	0.000	15.2	0.019	29.9	0.000	30.5	0.000	36.3	0.000	1.8	0.939
L.R. test (rounds 8 on)	59.2	0.000	7.3	0.290	20.4	0.002	4.2	0.651	4.6	0.602	1.1e-013	1.0000

Note: Estimates under columns “Uncons.” are unconstrained estimates of the joint probability of play. Estimates under columns “e-CE” are estimates of the joint probability of play constrained to satisfy correlated equilibrium restrictions with a level of tolerance of 0.05.

	Game 1			Games 2			Games 3		
Propensities	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
$(AABBBB, ABB)$									
Round of play	-0.207	0.144	0.151	-0.023	0.095	0.808	0.111	0.069	0.109
Constant	-0.606	1.108	0.585	-2.158	0.920	0.019	-1.801	0.739	0.015
$(BBBBBB, BBB)$ in Game 1, $(BBBBAA, BBA)$ in Games 2&3									
Round of play	-0.511	0.081	0.000	-0.233	0.038	0.000	-0.577	0.108	0.000
Constant	3.167	0.537	0.000	1.960	0.306	0.000	2.806	0.592	0.000
$(BABAAA, BBA)$									
Round of play	-7.610	1.777	0.000	0.245	0.074	0.001	0.242	0.108	0.025
Constant	-4.641	13.719	0.735	-3.732	0.858	0.000	-4.199	1.284	0.001
$(BABABA, ABB)$									
Round of play	-1.826	1765.379	0.999	0.207	0.191	0.279	0.400	0.241	0.097
Constant	-20.386	2531.606	0.994	-4.890	2.352	0.038	-6.874	3.209	0.032
$(BABBBA, ABB)$									
Round of play	0.617	465.235	0.999	-0.120	0.086	0.159	-0.261	0.080	0.001
Constant	-32.537	5690.214	0.995	-0.869	0.706	0.218	1.015	0.590	0.086
ω									
Round of play	-0.052	0.029	0.073	-0.061	0.013	0.000	-0.053	0.021	0.011
Constant	-1.185	0.207	0.000	-1.353	0.101	0.000	-1.938	0.175	0.000
Observations		171			375			180	
Log Likelihood		-817.5			-1725.0			-688.8	

Table 6: Noisy communication and play model

	Game 1			Games 2			Games 3		
Message, Play	Coef.	s.e.	p-value	Coef.	s.e.	p-value	Coef.	s.e.	p-value
				At round 5					
$(AAAAAA, AAA)$	0.329	0.051	0.000	0.274	0.031	0.000	0.331	0.053	0.000
$(AABBBB, ABB)$	0.064	0.032	0.048	0.028	0.014	0.042	0.095	0.035	0.006
$(BBBBBB/AA, BBB/A)$	0.607	0.055	0.000	0.607	0.035	0.000	0.306	0.054	0.000
$(BABAAA, BBA)$	0.000	-	-	0.022	0.011	0.041	0.017	0.013	0.183
$(BABABA, ABB)$	0.000	-	-	0.006	0.008	0.484	0.003	0.005	0.622
$(BABBBA, ABB)$	0.000	-	-	0.063	0.022	0.004	0.248	0.049	0.000
				At round 10					
$(AAAAAA, AAA)$	0.825	0.047	0.000	0.446	0.035	0.000	0.505	0.049	0.000
$(AABBBB, ABB)$	0.057	0.035	0.107	0.041	0.015	0.005	0.253	0.043	0.000
$(BBBBBB/AA, BBB/A)$	0.118	0.040	0.003	0.307	0.035	0.000	0.026	0.016	0.102
$(BABAAA, BBA)$	0.000	-	-	0.123	0.025	0.000	0.085	0.028	0.002
$(BABABA, ABB)$	0.000	-	-	0.027	0.017	0.121	0.028	0.025	0.262
$(BABBBA, ABB)$	0.000	-	-	0.056	0.021	0.008	0.103	0.035	0.003
				At round 15					
$(AAAAAA, AAA)$	0.966	0.030	0.000	0.404	0.068	0.000	0.343	0.079	0.000
$(AABBBB, ABB)$	0.024	0.029	0.416	0.033	0.022	0.126	0.299	0.080	0.000
$(BBBBBB/AA, BBB/A)$	0.011	0.008	0.179	0.087	0.026	0.001	0.001	0.001	0.385
$(BABAAA, BBA)$	0.000	-	-	0.379	0.082	0.000	0.195	0.074	0.008
$(BABABA, ABB)$	0.000	-	-	0.068	0.051	0.178	0.142	0.078	0.068
$(BABBBA, ABB)$	0.000	-	-	0.028	0.020	0.164	0.019	0.014	0.165

Table 7: Predicted Probabilities