An Experiment on Prisoner's Dilemma with Confirmed Proposals

Giuseppe Attanasi [*] – *Toulouse School of Economics (LERNA), France*

Aurora García Gallego[†] – *LEE, Universitat Jaume I, Castellón, Spain*

Nikolaos Georgantzís[†] – LEE-UJI, Castellón and *GLOBE*, *Universidad de Granada, Spain*

Aldo Montesano – *Bocconi University, Milan, Italy*

December 2011

Abstract

We apply an alternating proposals protocol with a confirmation stage as a way of solving a Prisoner's Dilemma game. We interpret players' proposals and (no) confirmation of outcomes of the game as a tacit communication device. The protocol leads to unprecedented high levels of cooperation in the laboratory. Assigning the power of confirmation to one of the two players alone, rather than alternating the role of a leader significantly increases the probability of signing a cooperative agreement in the first bargaining period. We interpret pre-agreement strategies as tacit messages on players' willingness to cooperate and on their beliefs about the others' type.

JEL classification: C72; C91; C92.

Keywords: Prisoner's Dilemma; Bargaining; Confirmed Proposals; Confirmed Agreement; Tacit Communication.

---

[*] Corresponding author: 21 allée de Brienne, 31000 Toulouse. Email: attanasi@cict.fr. Phone: +33 (0)5 61 63 57 76. Fax: +33 (0)5 61 12 85 20. Financial support by the Chair "Georges Meyer" in Mathematical Economics at Jean-Jacques Laffont Foundation (TSE) and by the French National Research Agency (ANR) through the project on "Fair Environmental Policies" is gratefully acknowledged.

[†] Financial support by the Spanish Ministry of Education (ECO2008-04636/ECON) and Junta de Andalucía (P07-SEJ-03155) is gratefully acknowledged.

1. Introduction

Since over half a century ago, a Prisoner's Dilemma is perceived as a metaphor of how the use of uncoordinated selfish actions traps groups of people into a non-cooperative equilibrium as opposed to the maximization of their collective welfare. Naturally, identifying the conditions under which the dilemma disappears in order for social optimality to be reached by uncoordinated individual actions has been a major theme of a prolific research agenda. Among other things, it has been established by now that repetition of a Prisoner's Dilemma favors the emergence of cooperation towards the collectively optimal outcome, due to a variety of reasons, like learning or, simply, as the subgame perfect equilibrium[1] of the supergame. Following these results, the likelihood of being re-matched with the same person in a social group also increases the probability of reaching the cooperative outcome.[2] Furthermore, evolutionary game theory has offered the theoretical background for a similar result, according to which cooperation is fostered as a collectively successful strategy destined to survive in the population even under relatively hostile conditions.[3] It is beyond the scope of this paper to exhaustively list the large number of rules and conditions that foster cooperation in a social dilemma.[4]

Interestingly, based more on experimental results than theory, we know that explicit communication drastically increases the ability of players to cooperate. A rather extreme design simultaneously introducing repetition, choice of partners with whom to be matched and uncontrolled verbal communication among players is reported in a paper by Tullock (1999), achieving almost full cooperation. However, verbal communication consists of a large set of potential and actual

---

[1] The paper by Selten and Stoecker (1986) presents a design aimed at disentangling these two effects of repetition on cooperation in a Prisoner's Dilemma.

[2] Andreoni and Miller (1993), Yang *et al*. (2007) and more recently Duffy and Ochs (2009) examine the effects of different aspects of the matching or re-matching protocol on the level of cooperation.

[3] For example, McNamara *et al*. (2004) introduce randomness in the variability of outcomes and show that it favors the survival of cooperation in the long run.

[4] Ranging from individual idiosyncratic factors like a subject's risk attitude (Sabater-Grande and Georgantzís, 2002), to social rules like the availability of punishment strategies against defectors (Wua *et al*., 2009; Dreber *et al*., 2008; with opposite findings), transfers among players (Charness *et al*., 2007) or sequentiality of players' moves (Clark and Sefton, 2001; Ahn *et al*., 2007).

messages and interchange protocols whose efficiency cannot be assessed, unless it can be isolated from other concurring factors.

In this paper, we formally prove and experimentally show that a simple language whose only messages are the possible strategies of the dilemma is sufficient for the cooperative outcome to be achieved. In fact, the implementation of such communication in the form of sequential bargaining with alternating proposals on the strategy to be played has an unprecedented cooperation-inducing effect on actual play.

Bargaining plays a central role in situations of interaction among economic agents. Since the seminal contributions by Nash (1950, 1953), bargaining is a central theme for research undertaken in the framework of cooperative and non-cooperative game theory. Furthermore, there is a huge literature on rationally justifiable play leading to cooperative outcomes in non-cooperative games.[5]

Several authors[6] have contributed to our understanding of the consequences of bargaining for the split of wealth among negotiating agents. In particular, Rubinstein's (1982) model illustrates an intuitively plausible and theoretically appealing way of reaching an agreement through sequential non-cooperative play. While the model has been criticized for a variety of reasons, there is hardly any doubt that it expresses most researchers' point of view on how bargaining should be modeled and on how it actually takes place if the negotiating parties have the right to make proposals as well as to reject those received by others in order to make their own counterproposals until an agreement is finally reached. The consensus on the plausibility of this bargaining protocol is compatible with the fact that bargaining models have been thought as stylized analogues of real world situations in which the negotiators aim at reaching an agreement concerning the distribution of wealth. However, in many occasions, bargaining processes pursue more complex objectives as compared to the split of a pie. Hence, the need for a more flexible accounting emerges, especially when dealing with social dilemmas.

---

[5] Relevant references are Harsanyi (1961), Friedman (1971), Smale (1980), Cubitt and Sugden (1994) and Innocenti (2008).

[6] Among these works, it is worth mentioning Harsanyi (1956, 1962), Sutton (1986) and Binmore (1987).

In this paper we provide a general bargaining process leading two players to sign a contract about how they will play in a specific non-cooperative game, namely the Prisoner's Dilemma. Suppose that two players bargain over the strategy profile to play in a Prisoner's Dilemma, given that each player knows the opponent's set of possible strategies. Then, there is an original social dilemma whose execution leads to the two players' final payoffs and a dynamic supergame whose actions in each bargaining period are *proposals* of strategies for the original game. The supergame ends when one of the two players confirms his/her proposal and the proposal of his/her opponent: the Prisoner's Dilemma is finally played according to the proposed and confirmed strategy profile.

The critical reader may observe that such situations are not fundamentally different from bargaining over the split of a pie, as long as they ultimately affect the distribution of wealth. From a technical point of view however, the basic difference between our framework and that of bargaining over the split of a pie is that, in our model, two agents bargain over their strategies in a 2x2 Prisoner's Dilemma. Apart from the obvious departure from Rubinstein's (1982) model due to the finiteness of the set of possible agreements,[7] in our setup, a confirmed agreement between bargaining agents concerns the pair of independent strategies in the original non-cooperative game.

Although we deal with an original one-stage game with a finite strategy space, the bargaining supergame built on it has potentially an infinite number of stages and involves an infinite number of strategies. Nonetheless, we show that, under mild assumptions, the equilibrium outcome of the bargaining process can be unique. The supergame ends when the agreement reached in two subsequent stages is confirmed by one of the two players. We call *equilibrium confirmed agreement* the corresponding equilibrium contract. When players alternate in exerting the power to end the game, the unique equilibrium confirmed agreement is the cooperative outcome of the Prisoner's Dilemma. This theoretical result does not hold when only one of the two players has the power to end the game, although the cooperative outcome is still an equilibrium confirmed agreement. For

[7] For a formal treatment of this issue see the insightful analysis by Muthoo (1991).

both supergames, we do not have any specific prediction about the timing of the equilibrium confirmed agreement.

We have tested our theory in the lab. Our experimental results provide strong support for the prediction of cooperation in social dilemma games with confirmed proposals. This contrasts with the moderate fit obtained from other sequential bargaining experiments testing perturbations of the standard alternating proposals framework.[8] In addition, we find that in the Prisoner's Dilemma modeled as a game with confirmed proposals, the fact that only one of the two players is exogenously given the power to end the game does not significantly affect the frequency of cooperation. On the contrary, the existence of an exogenous leader increases the likelihood of immediate cooperation, although it increases the average length of the negotiation needed to reach an agreement. Finally, our experimental results show how bargaining over the strategy to play in a Prisoner's Dilemma can be interpreted as a tacit communication device. In particular, we show that a glossary of bargaining semantics can be created in order to account for agents' implicit dialogues and signals contained in their proposals and confirmation strategies.

The remaining part of the paper is structured as follows. Section 2 describes a theoretical framework for the study of bargaining as a solution of two-players non-cooperative games. This theoretical framework is applied to the Prisoner's Dilemma in order to find the equilibrium confirmed agreements in two comparative bargaining settings. Section 3 describes the experimental design implemented to test the predictions of our theory and discusses the results. The observed bargaining dynamics in the Prisoner's Dilemma with confirmed proposals is interpreted in terms of basic heuristics providing a rationale for reaching an agreement on the cooperative outcome through recurrent strings of dialogs. Section 4 concludes.

[8] A prominent example is Binmore *et al*. (2007) including Rubinstein's reaction to the experimental test of his model, surprisingly arguing that his model should not necessarily be taken as a testable theory, but rather as a fable on human behavior. Earlier bargaining experiments include Sutton (1987), Neelin *et al*. (1988), Ochs and Roth (1989), Weg and Zwick (1999).

2. Bargaining Games with Confirmed Proposals

2.1 The rules of the bargaining game

Games with confirmed proposals are interactive strategic situations in which at least one player, in order to give official acceptance of a contract, must confirm his/her proposal in combination to the proposal of his/her opponent. The bargaining game can be a game with perfect or imperfect information and/or with complete or incomplete information. When information is incomplete, players can exploit the bargaining process to extract information on their opponent's type through their proposals.

Throughout the paper, we will assume that only two players are involved in the bargaining game. Let us denote by S_i the (finite) strategy space for player $i = 1,2$ in the *original game*. The (*super*-) *game with confirmed proposals* (henceforth GCP) has a potentially infinite sequence of bargaining *periods* t, with $t = 1, 2, \ldots, +\infty$. In each period, players are randomly assigned one over two roles: *proposer* or *responder*. Each period t is constituted by three *stages*, namely (t.I), (t.II), and (t.III). Only one player is active at each stage: at (t.I) only the proposer is active, at (t.II) only the responder, at (t.III) only the proposer. In the first two stages, (t.I) and (t.II), proposals of strategies are made; (t.III) is the confirmation stage: the active player decides whether to confirm or not the strategy profile that was proposed in the previous two stages. Therefore, only the player being proposer in a period can end the game at that period. The game ends when the proposer in a period confirms the strategy profile that was proposed in the first two stages of that period, leading to a *confirmed agreement* between the two players.

The proposer in each period t can be randomly chosen or picked-up according to a predetermined rule. In this paper, we focus on two limit cases where the proposer in each period is randomly chosen at the beginning of the game:

- GCP with *alternating* (*symmetric*) power of confirmation: once a player is randomly selected to be the proposer (responder) at the beginning of the game, he/she will play as proposer (responder) in period 1 and in each *odd* period; the opponent will play as proposer (responder) in each even period. Hence, players alternate in exerting the power to end the game (by confirming the agreement reached in a period). In this case, the rules of the game are symmetric, apart from the random selection of the proposer in the first bargaining period.

- GCP with *unilateral* (*asymmetric*) power of confirmation: once a player is randomly selected as proposer (responder) at the beginning of the game, he/she plays as proposer (responder) in *all* periods. Hence, the proposer in each bargaining period is always the same player: only this player has the power to end the game (by confirming an agreement reached in a period). In this case, each player has always the same role (proposer or responder) in each period of the game.

Suppose that player *i* starts the bargaining *super-game*, i.e. *i* and *-i* are respectively proposer and responder in period 1.

Stage (1.I) Player *i* communicates to player *-i* his/her will to follow a certain strategy $s_i^1 \in S_i$ in the original game. It means that *i* would follow s_i^1 if (and only if) the bargaining process would come to a confirmed agreement (the contract is subscribed) and s_i^1 would be part of this agreement.

Stage (1.II) Player *-i* replies to *i*'s proposal by communicating his will to follow strategy $s_{-i}^1 \in S_{-i}$ in the original game if (and only if) *i* will confirm his/her previous strategy s_i^1.

Stage (1.III) Player *i* can *c*onfirm or not the preceding strategy profile, i.e. c_i^1 can be *Yes* or *No*. If he/she *agrees* (c_i^1 = *Yes*) with the strategy profile (s_i^1, s_{-i}^1), he/she communicates to player *-i* that he/she *confirms* his/her proposal (i.e., he/she confirms that he/she will follow s_i^1 in the original game) knowing that player *-i* will follow s_{-i}^1 in the original game. In that case, the bargaining process ends in period 1 with a so-called confirmed agreement and the two players receive the payoffs corresponding to the strategy profile (s_i^1, s_{-i}^1) in the original game. If he/she *does not agree* (c_i^1 = *No*) with the strategy profile (s_i^1, s_{-i}^1), the two players move to the next period.

If the GCP is with *alternating* power of confirmation, player -*i* becomes proposer in period 2. In stage (2.I) he/she proposes a strategy $s_{-i}^2 \in S_{-i}$, which *could* be the same proposed in stage (1.II) or a different one, and so on and so forth.

If the GCP is with *unilateral* power of confirmation, player *i* is again proposer in period 2. In stage (2.I) he/she proposes a strategy $s_i^2 \in S_i$, which *could* be the same proposed in stage (1.I) and not confirmed in stage (1.III) or a different one, and so on and so forth.

In this paper we focus on GCP with no exogenous constraints on the proposal a player can make in two subsequent bargaining periods. Thus, the bargaining period $t + 1$ starts without any constraint imposed by 'proposal' – 'counterproposal' – 'no confirmation' in period t: for each $t = 1, 2, \ldots, +\infty$, and for each (s_i^t, s_{-i}^t) in t, *i*'s and -*i*'s set of feasible proposals in $t + 1$ are respectively the whole set S_i and the whole set S_{-i} both in the symmetric and in the asymmetric version of the game.[9]

2.2 Prisoner's Dilemma with Confirmed Proposals

Let us now analyze the GCP version of the most well-known social dilemma game, the Prisoner's Dilemma (henceforth PD). This means that the original game is a standard PD and the bargaining supergame built on it is an infinite dynamic game with perfect and complete information. The set of player *i*'s feasible proposals (coinciding with his/her actions in the original game) is $S_i = \{A, B\}$, for $i = 1, 2$. Figure 1 below shows, both the one-shot original game and, at the same time, all the possible outcomes of the bargaining (super)game with confirmed proposals built on it.

[**Figure 1** here]

The original game has only one Nash equilibrium in dominant strategies, the profile (A, A), leading to a non-efficient outcome. The same equilibrium outcome would be found in the standard two-

[9] Attanasi *et al*. (2011) provides a theoretical analysis of GCP with alternating power of confirmation where the bargaining periods are *overlapping*: in each bargaining period, the counterproposal of the second mover represents at the same time the proposal of the first mover in the subsequent bargaining period. Moreover, each time a player proposes the same strategy both in period t and in period $t + 1$, the game ends with a confirmed agreement in period $t + 1$, given that "re-proposal" leads to "confirmation". They call this mechanism "Games with *Chained* Confirmed Proposals".

stage game (without bargaining and without confirmation) in Figure 2, where player $i = 1, 2$ moves first and player -i observes his/her "proposal" before choosing his/her own.

[**Figure 2** here]

Let us now calculate the subgame perfect equilibrium outcome(s) of the GCP version of this game. Recall that we have defined as '*equilibrium confirmed agreement*' the subgame perfect equilibrium contract agreed upon by the two players, i.e. the proposal $a_i^t \in \{A, B\}$ and the counterproposal $a_{-i}^t \in \{A, B\}$ of the confirmed equilibrium outcome $(a_i^t,\ a_{-i}^t \mid a_i^t,\ Yes_i^t \mid (a_i^t, a_{-i}^t))$, given that i and -i play, respectively, as proposer and responder in the bargaining period t, in which the contract (a_i, a_{-i}) is confirmed.

The PD with confirmed proposals is represented in Figure 3. The payoff set of this supergame is the same as the original game in Figure 1, with the first of the two payoffs referring to player 1 and the second referring to player 2. The game in Figure 3 with $i = 1$ is a PD with *asymmetric* confirmed proposals (*unilateral* power of confirmation): the proposer in each bargaining period is always player 1 (randomly chosen at the beginning of the game). Hence, only this player has the power to end the game, by confirming one of the four possible agreements reached in a certain period. The game in Figure 3 with $i = 2$ is a PD with *symmetric* confirmed proposals (*alternating* power of confirmation): player 1 is randomly selected to be the proposer in period 1 and so also in each odd period; player 2 will play as proposer in each even period.

[**Figure 3** here]

In the GCP version of the PD, if the power of confirmation is *alternated* between the two players, then only the cooperative outcome can be confirmed in equilibrium, as formally stated below.

Proposition 1. The Prisoner's Dilemma with confirmed proposals and *symmetric* power of confirmation has a *unique* subgame perfect equilibrium *outcome*, the cooperative agreement, that can be confirmed in *any* bargaining period $t = 1, 2, \ldots, +\infty$.

Proof. Consider the game in Figure 3 with $i = 2$. First of all, notice that players cannot agree in equilibrium on the contract (B, A), giving the proposer a payoff of 5. In each bargaining period t, the proposer in that period will never confirm this contract, given that he/she can always commit to play the strategy (A, *Yes*), allowing his/her a payoff of at least 10 in period t. For the same reason, players cannot agree in equilibrium on the contract (A, B), giving the responder a payoff of 5. In each bargaining period t, the responder in that period will never reply to a A proposal with a B proposal: committing on replying with A in period t and then, becoming the proposer in $t + 1$, playing the strategy (A, *Yes*), allows him/her a payoff of at least 10 in period $t + 1$. Moreover, the contract (A, A) cannot be an equilibrium outcome. This can be verified by using a stationarity argument. Given that the game horizon is infinite, all subgames starting in even nodes are identical and the same holds for all subgames starting in odd nodes. Since the players are rational, strategy profiles confirmed in period t will be the same as the ones that would have been confirmed at $t + 2$, with $t = 1, 2, \ldots, +\infty$. Hence we can characterize a subgame perfect equilibrium based solely on stationary strategies. Suppose that (A, A, *Yes*) is an equilibrium outcome, leading to the payoff profile (10, 10) in period 1. In a stationary equilibrium, the payoff profile at the end of period $t = 1$ has to coincide with the payoff profile at the end of period $t > 1$, for each $t = 1, 2, \ldots, +\infty$. Therefore, given that the game starting in α and the one starting in δ are isomorphic (the set of strategies in the two games are the same and the original game is symmetric), we can assign to each non-terminal node at the end of bargaining period 1 (time δ) the payoff profile (10, 10). That would lead the first proposer, player 1, to choose *Yes* at the end of period 1 in every nodes apart from (B, A). In particular, since the payoff he/she obtains in the terminal node (A, A, *Yes*) is the same as in the non-terminal node (A, A, *No*), he/she is indifferent between confirming the contract (A, A) and not confirming it. Going backward, in any case the responder (player 2) would best-reply to A with A and to B with B. Hence, at the beginning of period 1, player 1 would propose B, player 2 responds with B and player 1 confirms, so leading to the payoff profile (15, 15) at the end of period 1, which contradicts that the confirmed agreement (A, A) is a stationary equilibrium outcome. Therefore, only

(*B*, *B*) can be an equilibrium outcome. Let us verify that the game ends in some period *t* with the following plan of actions: (*B*, *B*, *Yes*). Given that we assign to each non-terminal node at the end of bargaining period *t* the payoff profile (15, 15), the proposer does not confirm (*A*, *A*) and (*B*, *A*), confirms (*A*, *B*) and does not decline (*B*, *B*), given that she gets the same payoff in the terminal node (*B*, *B*, *Yes*) and in the non-terminal node (*B*, *B*, *No*). The payoff obtained by the proposer through confirming (*B*, *B*) in period *t* equals the highest payoff he/she can get by continuing the game; moreover, for any $\hat{t} > t$, this payoff can be obtained only by confirming the same agreement confirmed at *t*. Going backward, the responder would best-reply to *A* with *A* and to *B* with *A* or *B*, thus leading the proposer to confirm an agreement only in that period where she starts by proposing *B* and the responder replies to *B* with *B*. This agreement is (*B*, *B*).■

In the GCP version of the PD, if the power of confirmation is *unilateral*, then all outcomes but the worst one for the proposer can be confirmed in equilibrium, as formally stated below.

Proposition 2. The set of subgame perfect equilibrium outcomes in the Prisoner's Dilemma with confirmed proposals and *asymmetric* power of confirmation is $\{(A,A),(A,B),(B,B)\}$. Each of these *three outcomes* can be confirmed in *any* bargaining period $t = 1, 2, \ldots, +\infty$.

Proof. Consider the game in Figure 3 with *i* = 1. Notice that players cannot agree in equilibrium on the contract (*B*, *A*), giving the proposer a payoff of 5 (same argument used in the proof of Proposition 1). Players could instead agree in any period *t* in equilibrium on the contract (*A*, *B*), giving player 2 a payoff of 5. This is because player 2, being the responder in each bargaining period, cannot commit on any counterproposal allowing him/her a higher payoff. In fact, assigning to each non-terminal node at the end of bargaining period 1 the payoff profile (25, 5) leads player 1 to choose *No* at the end of period 1 in all but one node in which he/she is active. Indeed, given that player 1 gets the same payoff in the terminal node (*A*, *B*, *Yes*) and in the non-terminal node (*A*, *B*, *No*), according to the same weakly dominance criterion used in Proposition 1, he/she confirms (*A*, *B*). Going backward, the responder (player 2 in all periods) is indifferent between *A* and *B* in both nodes in which he/she is active. In the light of this, the proposer has a weak preference for starting with proposal *A* in *α*, in case he/she rationally attaches at least a small probability to player 2's counter-proposing *B* to *A*. Therefore, the agreement (*A*, *B*) can be confirmed in equilibrium in any bargaining period *t*. This result can be used to prove that also the contract (*A*, *A*) can be an equilibrium outcome in any period *t*. Suppose that when player 1 chooses *No* at the end of bargaining period 1, the payoff profiles in the continuation game are (25, 5) in correspondence to (*B*, *B*, *No*) and (10, 10) for all the other non-terminal histories. With this payoff structure of the continuation game, at the end of period 1, player 1 does not confirm either (*B*, *A*) or (*B*, *B*), confirms (*A*, *B*) and is indifferent between confirming or not (*A*, *A*). Going backward, the responder (player 2 in all periods) replies with *A* in both nodes in which he/she is active (ensuring herself a payoff of 10), hence leading player 1 to be indifferent between proposing *A* or *B* at the beginning of the period. Therefore, the agreement (*A*, *A*) can be confirmed in equilibrium in each bargaining period *t*. The fact that also the agreement (*B*, *B*) can be confirmed in equilibrium in any bargaining period *t* can be easily proved through the same stationarity argument used in the proof of Proposition 1. ■

Therefore, in the PD with alternating power of confirmation, the two players should at some point agree on the cooperative outcome. Moreover, in equilibrium the two players could cooperate immediately or they could need two, more or infinite periods to reach the cooperative outcome. Cooperation can be obtained in equilibrium also when the power of confirmation is unilateral. However, in the latter case, the cooperative agreement is not the unique equilibrium outcome of the bargaining supergame, given that all outcomes apart the worst one for the proposer can be sustained in equilibrium in each period. Therefore, the Nash equilibrium of the original PD is also an equilibrium outcome of its GCP version with unilateral power of confirmation. Conversely, if the power of confirmation is alternated, the Nash-outcome of the original PD cannot be confirmed in equilibrium in its GCP version.

3. An Experimental Prisoner's Dilemma with Confirmed Unconditional Proposals

3.1 Experimental Issues

Let us consider a strategic interactive situation with two players and with two strategies available to each one of them: *Defect* (A) and *Cooperate* (B). Assume also that the combinations of the two players' strategies yield a one-shot PD type of situation, as the one represented in Figure 1. Furthermore, assume that the two players bargain over which strategy profile they will play in the one-shot PD. Finally, suppose that they bargain using one of the two GCP protocols described in section 2.

Our first research question is: does the GCP version of the PD lead unambiguously to the cooperative outcome? From the theoretical analysis of section 2 we derive the hint that the power of confirmation should play a role in answering this question. Indeed, we have seen that the confirmed-proposal bargaining protocol always includes the cooperative Pareto-superior outcome among its subgame perfect equilibria, independent of whether the power of confirmation is symmetrically or asymmetrically allocated between the two players. However, the cooperative Pareto-superior outcome is the unique equilibrium agreement only if the power of confirmation is symmetrically distributed. Therefore, we expect the cooperative agreement to emerge more frequently in the symmetric GCP version of the PD, than in the asymmetric one.

Our second research question is: does the way in which the power of confirmation is distributed between the two players influence the speed at which an agreement is reached? Here we expect that a symmetric distribution of the power would reduce the number of bargaining periods, given that the equilibrium agreement is unique. However, recall that even with symmetric power of confirmation, "immediate" cooperation is only one of the possible equilibrium paths in the GCP. For example, in equilibrium it is possible that players propose and do not confirm for several times any among the contracts (A, A), (B, A), and (B, B), before finally confirming (B, B).

Our third research question is: can a GCP protocol lead to asymmetric gains/losses with respect to the (non-cooperative equilibrium) payoffs that players would obtain by playing the original PD? Among the two confirmed-proposal structures presented in section 2, only the asymmetric one embeds an equilibrium agreement where the proposer obtains his/her highest payoff possible and the responder accepts his/her lowest payoff possible. Conversely, the protocol in which both players can alternatively make a proposal and confirm it leads to exclude asymmetric agreements from the set of equilibrium confirmed agreements: both players should obtain an equal Pareto-superior gain by bargaining on which strategy to play in the original game rather than playing it one-shot.

These three research questions would shed some light on which features among those of the two suggested GCP structures are crucial for the possibly immediate and high level of cooperation. Suppose that, from an experimental point of view, we find the same (high) frequency of cooperative agreements confirmed both in the asymmetric environment and in the symmetric one, reached after few bargaining periods. This would lead us to state that alternated confirmation power, which is crucial in Rubinstein (1982), is not a necessary condition for bargainers' cooperation in the PD with confirmed proposals. At that point, we could conclude that the key feature for cooperation in the PD is the mechanism of 'proposal' – 'counterproposal' – '(no) confirmation' itself.

Finally, we are interested in GCP as a form of communication between players involved in social dilemmas. Through the bargaining structure developed, we believe that it could be understood how players communicate in social dilemmas without using 'explicit' communication devices. In particular, we could shed some light on what a subject *wants* to communicate to another, on what he/she *is able* to communicate as well as to understand when the 'proposal' – 'counterproposal' – '(no) confirmation' mechanism is used to "solve" a social dilemma.

3.2 Experimental design

Experimental subjects were voluntary undergraduate students in Economics recruited at the Laboratory of Experimental Economics at *Universitat Jaume I* in Castellón (Spain), and at *Bocconi*

University in *Milan* (Italy). Sessions were conducted in appropriate rooms where subjects were seated in isolated cubicles in front of computer terminals that were connected through a computer network. A total of 216 experimental subjects (91 women, 125 men, average age = 21.8 years, age range: 19–25 years) participated in our experiments, with each subject participating only once. Average earnings were approximately 15 euro per subject. The experiment was programmed using the z-Tree software (Fischbacher, 2007). Two treatments were run, both being built starting from the one-shot PD depicted in Figure 1. The same number of subject pairs ($N = 54$) has participated in each one of the two treatments. The first treatment was the *Sym*metric GCP version of the game and the second was the GCP with *Asym*metric power of confirmation, denoted respectively by GCP-*Sym* and GCP-*Asym*. Subjects were informed that each session would last (as specified also in the recruitment wall) a maximum of 2 hours. Also, in order to avoid the formation of uncontrolled beliefs regarding the implications of this time limit for the number of periods that would be played, we limited by design the maximum time that each negotiation period could last up to 3 minutes (80 seconds per subject's proposal and 20 seconds for the confirmation stage). This guaranteed the subjects that at least 40 periods could be played even in the case of the longest possible pre-agreement process. If no agreement were reached within the 2 hours limit, subjects would receive the minimum payoff of 5 euro. In fact such a disagreement payoff was never used. Furthermore, average bargaining period time was eventually lower than 2 minutes, whereas the longest (in periods and real time) negotiating process ended within an agreement in the 65th period, after 1 hour and 47 minutes (average bargaining period time slightly lower than 100 seconds). However, this happened in a small number of rather extreme cases, given that in the majority of cases an agreement was reached in less than 10 minutes. Let us analyze the specific features of each of the two treatments. In both treatments, at the beginning of the experimental session, pairs are randomly formed and are fixed during the whole session. Players belonging to the same pair are sited in different computer rooms. Within each pair, each player is randomly chosen to play either the role of proposer or responder. More specifically:

- **Symmetric GCP Treatment** (GCP-*Sym*). Within each pair, each player is randomly selected to play the role of a proposer or a responder in the first bargaining period. Roles change every period, that is, within each pair, each time the proposer in period t does not confirm the strategy profile proposed, he/she plays as a responder in period t +1, and vice versa. Hence, the two players alternate in exerting their power to end the game. Proposals in a bargaining period are not conditional to those made in the previous period: each time a period ends without confirmation, the proposer in the new period makes a proposal which can be the same or different to the one his/her opponent made - and not confirmed - in the previous period. The game ends when one of the two players decides to confirm the sequence proposal-counterproposal within a period in which he/she has the role of proposer. Figure 3, with i = 2, depicts the GCP proposed to subjects participating in this treatment.
- **Asymmetric GCP Treatment** (GCP-*Asym*). Within each pair, each player is randomly selected to play the role of a proposer or a responder for the whole game (roles are fixed during the whole session). Therefore, if the proposer decides not to confirm his/her proposal (once known his/her opponent's choice) in a bargaining period, he/she starts the next period by making a new proposal, which can be the same or different to the one made - and not confirmed - in the previous period. The game ends when the player being assigned the role of proposer decides to confirm the sequence proposal-counterproposal within a bargaining period. Figure 3, with i = 1, depicts the GCP proposed to subjects participating in this treatment.

In both treatments, subjects in a pair do not have to wait for the other pairs to end the game. Once a pair reaches an agreement, the two players belonging to this pair leave their cubicles, and proceed to a separate room in which they are individually paid.

3.3 Experimental Results

The same number of subject pairs (N = 54) has participated in each one of the two treatments. In this section we discuss the facts obtained from the statistical analysis of our data. For each treatment, half of the sessions were run in Castellón (Spain) and the other half in Milan (Italy): we find no significant difference in behavior between pairs participating in sessions run in the former location and those participating in sessions run in the latter location. In fact, similar frequencies of specific patterns of proposals are found in each location.

The simplest way of looking at our data is by observing the frequency of cooperation and the speed of reaching the corresponding agreement. Table 1 informs us on the first of these two issues. Almost all pairs reach the cooperative confirmed agreement in the two GCP treatments. Specifically, the cooperation ratio is around 91% in the symmetric treatment and around 93% in the asymmetric treatment.

[**Table 1** here]

In both treatments, only 2 pairs over 54 (3.5%) reach a confirmed agreement to behave à la Nash, (*A*, *A*), in the Prisoner's Dilemma. In one of the two cases in which this occurred in treatment GCP-*Asym*, this happened in the first period while in the second case it happened after 33 bargaining periods. In both cases (see Appendix 1), the proposer starts the first bargaining period by playing the non-cooperative action and the responder plays the non-cooperative action in any bargaining period. In fact, in the first case, the proposer plays the non-cooperative action, and is imitated by the responder. Then, the proposer confirms. Hence, it seems as if the Nash outcome was imposed by the proposer, given that he/she does not try any cooperation at all. In the second case, the proposer plays the non-cooperative action 33% of the times and the cooperative one 66% of the times. However, independently of the proposal, the responder replies with the non-cooperative action. Hence, it seems as if the Nash equilibrium here was imposed by the responder. It is striking that in the asymmetric treatment the non-cooperative Nash outcome was observed only in the cases in which one of the two players was absolutely committed to non-cooperative behavior.

Similarly, the asymmetric outcome involving the proposer defecting while the responder cooperates appeared in another two occasions in each treatment. In both cases in which this occurred in treatment GCP-*Asym*, this agreement was reached in the first bargaining period (see Appendix 1), while in treatment GCP-*Sym* this agreement was reached in one pair in the first bargaining period and in the other one in the second bargaining period (see Appendix 2). In all cases where the (*A*, *B*) contracts was signed, during their payment by the experimenters, the four involved responders recognized that they had accepted the lowest possible payoff by mistake. The same happened also to the only proposer who confirmed the (*B*, *A*) agreement in treatment GCP-*Sym*.

We have spent these first lines of the discussion of our results on the few observations deviating from the cooperative outcome. As predicted by our theoretical analysis, the cooperation obtained in the symmetric treatment is pervasive. Furthermore, the cooperative outcome has received similarly ample support in the asymmetric treatment. However, we are also interested in the way in which the cooperative outcome is achieved. We analyze first the length of the bargaining process in the two treatments. In Figure 4, we represent the relative frequencies of the *agreement period*, i.e. the period in which the confirmed agreement was signed in each pair, disentangled by treatment.

[**Figure 4** here]

First of all, it is worth observing that, despite the almost[10] infinite nature of the GCP structure, in the vast majority of cases, cooperation is reached in few periods. However, there seems to be some differences in the timing of agreements obtained under the two protocols, the symmetric and the asymmetric one. The average length of a bargaining finishing with a cooperative agreement is only 3.5 periods in treatment GCP-*Sym*, while it is 9 periods in treatment GCP-*Asym*. Nevertheless, the modal length of a bargaining leading to cooperation is 3 periods in GCP-*Sym*, while it is only 1 period in GCP-*Asym*. In particular, considering only subjects' pairs who have agreed on signing the (*B*, *B*) contract, in the symmetric treatment only 6 over 54 pairs (11.1%) confirm the cooperative agreement in the first bargaining period, while in the asymmetric treatment immediate cooperation

[10] Recall that our experimental subjects agreed in advance on a 2 hours maximum staying in the lab, hence the GCP that we tested experimentally cannot be exactly classified as "infinite". However, the exogenously imposed short duration of each bargaining period and the high speed of bargaining (25% of bargaining periods lasted less than 30 seconds) have led subjects to think that a potentially huge number of negotiations was possible. Indeed, in a final questionnaire subjects have been asked to state how many bargaining periods they could have played in case of no confirmation in their pair (that is instead what happened) and the average answer has been 93. Notice that the distribution of answers to this question is not correlated with the agreement periods of the pairs.

comes out in 17 pairs over 54 (31.5%). Therefore, the existence of an exogenous leader almost triples the likelihood of an immediate cooperative agreement in the Prisoner's Dilemma with confirmed proposals.

The most interesting difference between the two treatments can be isolated once we look at the sequence following an initial cooperative proposal. While the number of initial cooperative proposals has been very similar across treatments, the frequency of cooperative replies to them has been different. Specifically, in treatment GCP-*Asym* more than a third (37%) of proposers (20 pairs) starts with a cooperative strategy which is also adopted immediately by most (18) responders and the cooperative outcome is almost globally confirmed (17 over 18). In any case, in the remaining 3 pairs with initial cooperative proposals, the cooperative agreement is finally confirmed by bargaining period 3 at most. Surprisingly enough, in treatment GCP-*Sym* the number of first-period proposers starting with a cooperative proposal does not significantly differ from that of treatment GCP-*Asym* (19 vs 20). However, few of the corresponding responders reciprocate with cooperation (6 over 19). Nonetheless, even in this treatment, first-period cooperative proposers receiving a cooperative reply immediately confirm cooperation (6 over 6).

Therefore, the lack of immediate cooperation in the symmetric treatment seems to be due to the first-period responders' high willingness to move to the second bargaining period, where they will "become" proposers, hence being assigned the power of confirmation at least once. In fact, in 9 over the 13 pairs where the first-period cooperative proposer is not reciprocated with cooperation by the first-period responder, the cooperative agreement is finally confirmed by bargaining period 3 at most. These 9 respondents "experience" the pleasure of the power of confirmation once and then they suddenly sign a cooperative agreement.

These data also show that the initial willingness to cooperate of first-period proposers is able to minimize the difference in the "proposer's power effect" between the two treatments. This point requires a more thorough discussion. In treatment GCP-*Asym*, being randomly selected as proposer means becoming the only player in the pair who will decide when the negotiation will end.

Conversely, in treatment GCP-*Sym*, this "power" lasts only one period before it is assigned to the opponent in the pair. Despite that, when the first-period proposer starts by proposing the cooperative strategy *B* in the first bargaining period, it seems that he/she commits to this proposal independent of the treatment. In fact, among the 20 proposers who start with *B* the first bargaining period in treatment GCP-*Asym*, there are 18 (90%) who behave cooperatively during the whole negotiation. This means that they always propose *B* in each bargaining period and they confirm a (*B*, *B*) agreement the first time where it is possible. A similar result can be stated for treatment GCP-*Sym*: among the 19 first-period proposers who start by proposing strategy *B* in the first bargaining period, there are 17 (89.5%) who behave cooperatively during the whole negotiation. This means that, in addition to always proposing *B* and always confirming (*B*, *B*) in periods where they play as proposers (odd periods), they behave cooperatively also when they "become" respondents (even periods): they always reply with *B* in each period where they are respondents and they receive a *B* proposal by their opponent.

Therefore, we can conclude that whenever a subject wishes to reach the cooperative agreement very soon, being assigned the role of proposer for the whole bargaining game (GCP-*Asym*) or just for the first bargaining period (GCP-*Sym*) does not make any difference. In other words, proposing a cooperative strategy in the first bargaining period is a kind of commitment to the cooperative outcome for the first-period proposers, both in the asymmetric and in the symmetric treatment.

Finally, it is worth mentioning that in our experimental sessions, even when the first-period proposer started with the non-cooperative strategy *A* in the first bargaining period, the alternating proposals protocol GCP-*Sym* has never led to a bargaining process lasting more than 10 periods. Conversely, with proposers starting with strategy *A*, the unilateral proposals protocol GCP-*Asym* has produced some very long negotiations, including a small percentage of cases (3 over 54) in which bargaining lasted for over 60 periods. In this three cases cooperation was arguably forced by the end-session alert.

In summary, the existence of asymmetric confirmation power has a dramatic positive effect on the frequency of immediate cooperation agreements, but it also entails some low risk of extremely long bargaining processes.

3.4 Beyond GCP testing: Proposal/confirmation strings as bargaining semantics

In order to give a complete explanation of all the patterns and dialogues emerging from the experimental data of the two GCP protocols, let us consider first different types of subjects, concerning factors like their understanding of the rules of the GCP, their preferences on outcomes, their bargaining toughness, as well as their beliefs on their opponents' personality.

First of all, given the simplicity of the bargaining game and the fact that our subjects' behavior complies with the predicted equilibria in the vast majority of the cases, we cannot reject common knowledge of the game within subject pairs.

Secondly, recall that in the theoretical model impatience is not an issue, although it may implicitly be present in our *weak dominance* assumption (see proof of Proposition 1 and of Proposition 2). But even in that case its presence alone and not the degree of it is required to break the tie between equal payoffs in different periods along the equilibrium path. Therefore, our assumption concerning time preferences regards an arbitrarily low impatience that makes subjects prefer an earlier to a later agreement when the payoffs achieved are identical across periods. However, our experimental results offer much stronger support for early-period cooperative agreements, compatible with strongly impatient behavior. An additional explanation for the observed behavior could be subjects' *toughness*, defined as their ability to postpone an agreement in order to achieve a better outcome. Tougher (softer) negotiators are more (less) willing to postpone an agreement in order to achieve a given payoff increase. This concept is sufficiently general to encompass several idiosyncratic differences between two players in a pair. For example, such a difference in players' toughness may be the result of players' expectations from the game, their beliefs on own and others' negotiating abilities or even beliefs concerning the relation between own and other players' patience.

Suppose that the game is with complete information and that both agents are self-interested and extremely tough. Assume also that they both believe that their opponent is self-interested and extremely tough. Thus, they should propose (*B*, *B*) in the first bargaining period, followed by a confirmation by the proposer. This is independently from the distribution of the power of confirmation. If this power is symmetric, we know that the cooperative outcome is the only equilibrium outcome in the general case. If this power is asymmetric, players' toughness is responsible for equilibrium selection.

However, players' toughness can be responsible for the speed of coordination among players too. Recall that the theoretical analysis of section 2 suggests in both protocols that the cooperative agreement could be achieved in equilibrium in every bargaining period. Our hint is that the tougher players are in negotiation, the smaller the number of bargaining periods required for them to coordinate. If (*B*, *B*, *Confirm*) is not obtained in the first bargaining period**,** it means that at least one of the previous hypotheses is not fully satisfied (e.g., one of the two players is not tough enough, or he/she thinks that the other is not tough enough). Nonetheless, obtaining (*B*, *B*, *Confirm*) in the first period does not necessarily mean that all hypotheses mentioned above are satisfied. For example, an irrational agent, choosing randomly, could propose the cooperative action and/or confirm the cooperative action profile. Our experimental data show that (*B*, *B*, *Confirm*) is not obtained in the first period for over 2/3 of all pairs.

The role of social preferences seems not so relevant in justifying this result. In fact, suppose that players in a pair can be:

- *self-interested*, i.e. for player 1 the payoff (25, 5) is better than (15, 15), which in turn is better than (10, 10), which in turn is better than (5, 25); for player 2 it is (5, 25) better than (15, 15), better than (10, 10), better than (25, 5);

- motivated (also) by ‘moderate’ self-interest and *fairness*, i.e. for player 1 it is (15, 15) better than (25, 5), better than (10, 10), better than (5, 25); for player 2 it is (15, 15) better than (5, 25), better than (10, 10), better than (25, 5);
- motivated (also) by *inequality aversion*, i.e. for player 1 it is (15, 15) better than (10,10), better than (25, 5), better than (5, 25); for player 2 it is (15, 15) better than (10, 10), better than (5, 25), better than (25, 5).

If both players are rational, tough and have perfect information and correct beliefs on their opponent, whatever combination of the three types of preferences above (i.e., 1 and 2 both self-interested, 1 and 2 both moderately altruist, 1 and 2 both inequality adverse, 1 inequality adverse and 2 altruist, and so on) leads again to the Pareto-efficient equilibrium outcome (*B*, *B*, *Confirm*) both in the symmetric confirmed-proposal treatment and in the asymmetric one.

Therefore, the reason of the continuation of the game after the first period for approximately 2/3 of all pairs has to be explained by weakening some other assumptions. The hypothesis that could be most easily weakened is the one about the beliefs on the opponents’ ability and/or willingness to play: at least one of the two players thinks that his/her opponent is not smart enough and/or not tough enough, etc.

A careful look at the strings of strategies obtained from our experiment reveals that all the dynamic patterns observed could be interpreted as dialogues between the two negotiating parties. The question we address in this section is how different types of “signals” can be sent by each player to his/her opponent in a cheap-talk bargaining context before one of them confirms a given strategy profile. In fact, we argue that in the specific case of bargaining about which strategy to play in a Prisoner’s Dilemma, both players aim at eliminating the asymmetric outcomes (*Cooperate*, *Defect*), and (*Defect*, *Cooperate*). Thus, it is of little if any relevance whether both players, or just one of them, have the right to confirm an announced strategy profile. For reasons which become clear by the end of the section, we concentrate on the case where the power of confirmation is asymmetric.

The bargaining semantics that we elaborate for the asymmetric treatment can be easily extended to the symmetric one.

The basic ingredients of the dialogues are questions and answers concerning each other's level of strategic competence or toughness. Therefore, as conductors of the information on other players' types, the fundamental questions and answers are specific to the roles of proposer and responder. As captured by the difference in first-period agreements across treatments, the value of such information seems to be recognized by our subjects. This explains the fact that in the symmetric treatment, first-period responders want to use their turn as proposers in order to obtain themselves information on their opponent's idiosyncrasy.

Not surprisingly, the most frequent dialogue was of the type – "Would you be nice to me if I were nice to you?" – "Yes" – "Then, let's cooperate", preceded or disturbed by some informative messages concerning subjects' strategic competence or toughness and the resulting possibilities of benefiting from others' "weaknesses".

Before we proceed with the analysis of observed bargaining patterns, we establish a glossary of 'proposal-response' *strings* with their corresponding verbal interpretations. The rationale behind the dialogues constructed below is founded on the following basic heuristics, which we claim are used by all pairs of humans who interact with each other in a strategic context like ours:

1. *Tough agents wish to know whether the others with whom they interact are tough*.
2. *If not, they take advantage of this*.
3. *Otherwise, they realize the benefits from cooperation, but they still fear that they may be fooled*.
4. *Once this fear is removed by the freedom of re-negotiation following unfair or inefficient outcomes, cooperation is the unique reachable agreement*.

In terms of the Prisoner's Dilemma context considered here, the following *strings of dialogues* provide an exhaustive list of basic conversations which can be used to understand the bargaining dynamics we have observed.

*S*1. *"A-B-Confirm":*

- "Are you tough?"

- "No, I am not".

- "Then, I'll take advantage of this".

It is straightforward that, in order to take advantage of the responder's "light" bargaining attitude, the "*A-B*" sub-string should be followed by an immediate confirmation by the proposer.

*S*2. *"A-A-Withdraw":*

- "Are you tough?"

- "Yes, I am".

- "Then, we can both do better than being competitive to each other".

It is also straightforward to see why the "*A-A*" sub-string will not be confirmed by the proposer, who realizes that he/she can do better than obtaining the non-cooperative payoff of the game. This is a very strong and clear-cut prediction in the symmetric GCP framework, because the Nash equilibrium of the non-cooperative game is ruled out as one of the least expected outcomes. However, it belongs to the set of equilibrium outcomes of the PD with asymmetric confirmed proposals. Nevertheless, in order to refer to a tiny percentage of outliers contradicting this prediction, we use the string *"A-A-Confirm"* denoted by *S*2-Nash.

Following the observation of the proposer's toughness, *S*2 could be followed either by a proposal to cooperate or by another *S*2 or by a whole series of them. In the latter case, we will refer to a series of *S*2 by the term *toughness challenge*, aimed at eliciting the responder's toughness. It is straightforward to see why a toughness challenge followed by a *S*1 string should be interpreted as the responder's lack of toughness or at least as a large difference in the two players' toughness. According to the same reasoning, the proposer's commitment to a very long series of *S*2 repetitions should be a result of his/her belief that the responder's toughness is significantly lower than his/her own.

As stated already, if the "*A-A*" sub-string belongs to a toughness challenge, it will not be confirmed. Rather, it will be followed by an invitation of the proposer to cooperate. However, the responder may now want to elicit the proposer's strategic competence or toughness, taking at the same time a 'revenge' on his/her rival's initial doubt concerning his/her own toughness. This will give rise to a third type of string:

*S*3-*A*. *"B-A-Withdraw"*:

- "Ok, then. Let's cooperate".
- "(Wait! It is my turn to know:) Are you smart?"
- "Yes I am!"

It is very unlikely that this string of dialogue will uncover the proposer's lack of strategic competence if a string like *S*2 has preceded *S*3, revealing the proposer's perfect understanding of the strategic situation. However, the *revenge motive* may still hold strong. In a similar manner as in the case of *S*2, a series of *S*3 may be observed corresponding to a toughness challenge by the responder. Also, the incentive of checking the responder's strategic competence or toughness may even emerge after both players proposing *B*. Then, in some occasions the proposer may first check the responder's willingness to cooperate, then ensure that the latter is a smart player. This will give rise to an alternative *S*3:

*S*3-*B*. *"B-B-Withdraw"*:

- "Ok, then. Let's cooperate".
- "Ok, then. Let's cooperate".
- "(Wait! It is my turn to know:) Are you smart?"

The predicted end point of all types of such dialogues within a pair of players being both strategic competent and tough enough will be:

*S*4. *"B-B-Confirm"*:

- "Ok, then. Let's cooperate".

- "Ok, then. Let's cooperate".

- "Confirmed".

This will be the end point of the bargaining process leading to the confirmation of the "*B-B*" sub-string of strategy proposals, yielding an agreement on the predicted strategy profile (*B*, *B*). It should be observed that the abstract setting of the Prisoner's Dilemma used in our experimental setting contains several of the aspects which are central in more generic bargaining situations involving payoff asymmetries. Those aspects should be relevant in the presence of fairness considerations and Pareto dominance, which should guide the agents' endeavors towards economic efficiency.

In Appendix 1 (treatment GCP-*Asym*) and Appendix 2 (treatment GCP-*Sym*) we report our experimental data (54 pairs per treatment) and we classify them according to short or longer dialogues consisting of these four strings. We identify 8 dialogues that provide an exhaustive list of the bargaining histories observed in the two GCP treatments. A significant part (1/3) of them totally coincides with *S4*, whereas over 90% of them ends with *S4*.

While we do not agree with the approach of calling subjects' mistakes those observations that contradict a theory, we feel that the ability of our setup to organize such a large percentage of our observations legitimates some final remarks on decision-making errors. The possibility of making mistakes, which can derive from the weakening of one of the hypotheses introduced in the beginning of this section, is the reason underlying the responder's behavior in dialogues 3 and 3-bis of the *GCP-Asym* treatment (see Appendix 1) and in dialogues 3 of the *GCP-Sym* (see Appendix 2). This possibility concerns both agents, although for the proposer only mistakes in confirmation are relevant.[11] For the responder, making a mistake in the proposal is crucial, given that he/she cannot confirm or withdraw it. Notice that when at least one subject in the pair is sure that he/she will never make mistakes but he/she believes that his/her opponent will make mistakes with positive probability, then the game could never end, even if both agents were extremely tough. However, as the play unfolds both players can signal, through their proposals, counterproposals, and withdrawals, that they are smart and though enough, thus influencing the opponent's belief or hope that they could make a mistake in a subsequent stage of the bargaining process. In other words, they have the means to convince the opponent that he/she holds a wrong belief about their probability of an erroneous decision. Once both players are sure that their opponent will not make a mistake and that he/she is tough enough, they both agree on a (*B*, *B*, *Confirm*) outcome, at some subsequent period of the bargaining game.

A skeptical reader may hurry to argue that the four strings of dialogues and their variations mentioned above are simply all possible combinations of the two players' strategies in the Prisoner's Dilemma with confirmed proposals. However, it should be noted that our analysis involves predictions on the timing of the strings in a given dialogue, on the outcome of the confirmation/withdrawal choice and even on the possible repetitions of each one of them before the end point, *S*4.

Contrary to explicit verbal communication, we claim that the '*bargaining semantics*' proposed here contain the necessary and sufficient syntaxes for a dialogue between bargaining agents in non-cooperative contexts as social dilemmas. Many authors have explored the role of communication on the ability of agents to reach cooperative outcomes. Experimentalists have often used open or controlled verbal protocols in order to establish the cooperation-enhancing potential of communication. Cheap-talk signaling of agents' cooperative intentions or publicizing subjects' preferences or belief elicitation results have also been studied and shown to enhance cooperation.

[11] Making a mistake in the proposal or in the withdrawal has the only consequence of sending an involuntary wrong signal to the responder. However, in Dialogue 5-bis we observe a proposer making a mistake in the confirmation, as recognized by himself/herself during his/her payment by the experimenters.

However, our approach achieves *full* cooperation in *all* cases in which *strategic competent and tough* players are involved. Our experimental results confirm that the confirmed-proposal framework offers the natural and minimal semantic bridge between cooperative and non-cooperative behavior, implementing cooperation in a Prisoner's Dilemma with at least 90% of success. From now on we know that the minimal semantic charge required for an individual's *vocabulary* to support non-cooperative bargaining leading to full cooperation in social dilemmas is as little as two sentences:

- *"Are you tough?"*
- *"Let's cooperate"*.

The remaining job is done by a context in which the proposer can withdraw or confirm a potentially cheap-talk string of signals into an actual strategy profile.

4. Conclusions and further applications

Throughout the paper, we have shown through a behavioral game-theoretical framework the positive effect of potentially binding messages (defined as confirmed proposals) on agents' ability of coordinating on Pareto-efficient outcomes not belonging to the set of equilibrium outcomes of the traditional Prisoner's Dilemma.

From an experimental point of view, this particular bargaining structure applied to social dilemma games seems to lead to *cooperation* more than do other cooperation-enhancing mechanisms.[12] In addition, in the Prisoner's Dilemma modeled as a Game with Confirmed Proposals, alternating *power of confirmation* does not significantly affect the frequency of cooperation. On the contrary, the existence of an exogenous leader increases the likelihood of immediate cooperation, although it entails some risk of very long negotiation games.

[12] For a recent example and overview of related references, see Nikiforakis *et al*. (2010).

Finally, our experimental results show how Games with Confirmed Proposals can be used to create a glossary of bargaining semantics for tacit communication among agents concerning their *strategic competence*, *toughness in bargaining*, *beliefs about opponents' idiosyncratic characteristics*, etc., through the signals contained in their proposals and confirmation strategies. We claim that the 'proposal' – 'counterproposal' – '(no) confirmation' mechanism is extremely successful in promoting cooperation under the opponent's type uncertainty that characterizes social dilemmas.

We feel that this finding needs further investigation in future research, because of its implications for the organization of bargaining processes aiming either at maximizing the likelihood of immediate cooperative agreements or at minimizing the average or the maximal time of the negotiations needed to "solve" a social dilemma.

References

Ahn, T. K., Lee, M. S., Rutan, L. and J. Walker, 2007. Asymmetric Payoffs in Simultaneous and Sequential Prisoner's Dilemma Games, *Public Choice*, 132, 353–366.

Andreoni, J. and J. H. Miller, 1993. Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence, *Economic Journal*, 103, 570–585.

Attanasi, G., García-Gallego, A., Georgantzís, N. and A. Montesano, 2011. Non Cooperative Games with Chained Confirmed Proposals, in L. A. Petrosjan and N. A. Zenkevitch (Eds.) *Contributions to Game Theory and Management IV*, 19–32.

Binmore, K., 1987. Perfect Equilibria in Bargaining Models, in K. Binmore and P. Dasgupta (Eds.), *Economics of Bargaining*, Cambridge University Press, Cambridge.

Binmore, K., Swierzbinski, J. and C. Tomlinson, 2007. An Experimental Test of Rubinstein's Bargaining Model, ELSE Working Paper no. 260.

Charness, G. B., Fréchette, G. and C-Z Qin, 2007. Endogenous Transfers in the Prisoner's Dilemma Game: An Experimental Test of Cooperation and Coordination, *Games and Economic Behavior*, 60, 287–306.

Clark, K. and M. Sefton, 2001. The Sequential Prisoner's Dilemma: Evidence on Reciprocal Altruism, *Economic Journal*, 111, 51–68.

Cubitt, R. and R. Sugden, 1994. Rationally Justifiable Play and the Theory of Non-cooperative Games, *Economic Journal*, 104, 798–803.

Duffy, J. and J. Ocks, 2009. Cooperative Behavior and the Frequency of Social Interaction, *Games and Economic Behavior*, 66, 785–812.

Dreber, A., Rand, D. G., Fudenberg, D. and M. A. Nowak, 2008. Winners don't Punish, *Nature*, 452**,** 348–351.

Fischbacher, U., 2007. Z-Tree: Zurich Toolbox for Readymade Economic Experiments, *Experimental Economics*, 10, 171–178.

Friedman, J. W., 1971. A Non-cooperative Equilibrium for Supergames, *Review of Economic Studies*, 38, 1–12.

Harsanyi, J., 1956. Approaches to the Bargaining Problem before and after the Theory of Games: A Critical Discussion of Zeuthen's, Hicks', and Nash's Theories, *Econometrica*, 24, 144–157.

Harsanyi, J., 1961. On the Rationality Postulates Underlying the Theory of Cooperative Games, *Journal of Conflict Resolution*, 5, 179–96.

Harsanyi, J., 1962. Bargaining in Ignorance of the Opponent's Utility Function, *Journal of Conflict Resolution*, 6, 29–38.

Innocenti, A., 2008. Linking Strategic Interaction and Bargaining Theory: The Harsanyi-Schelling Debate on the Axiom of Symmetry, *History of Political Economy*, 40, 111–132.

McNamara, J. M., Barta, Z. and A. I. Houston, 2004. Variation in Behaviour Promotes Cooperation in the Prisoner's Dilemma Game Nature, *Nature*, 428, 745–748.

Muthoo, A., 1991. A Note on Bargaining over a Finite Number of Feasible Agreements, *Economic Theory*, 1, 290–292.

Nash, J., 1950. The Bargaining Problem, *Econometrica*, 18, 155–162.

Nash, J., 1953. Two-person Cooperative Games, *Econometrica*, 21, 128–140.

Neelin, J., Sonnenschein, H. and M. Spiegel, 1988. A further Test of Noncooperative Bargaining Theory: Comment, *American Economic Review*, 78, 824–836.

Nikiforakis, N., Normann, H.-T. and B. Wallace, 2010. Asymmetric Enforcement of Cooperation in a Social Dilemma, *Southern Economic Journal*, 76, 638–659.

Ochs, J. and A. Roth, 1989. An Experimental Study of Sequential Bargaining, *American Economic Review*, 79, 355–384.

Rubinstein, A., 1982. Perfect Equilibrium in a Bargaining Model, *Econometrica*, 50, 97–109.

Sabater-Grande, G. and N. Georgantzís, 2002. Accounting for Risk Aversion in Repeated Prisoners' Dilemma Games: An Experimental Test, *Journal of Economic Behavior and Organization*, 48, 37–50.

Selten, R., and R. Stoecker, 1986. End Behavior in Finite Prisoner's Dilemma Supergames, *Journal of Economic Behavior and Organization*, 7, 47–70.

Smale, S., 1980. The Prisoner's Dilemma and Dynamical Systems Associated to Non-cooperative Games, *Econometrica*, 48, 1617–1634.

Sutton, J., 1986. Non-cooperative Bargaining Theory: An Introduction, *Review of Economic Studies*, 53, 709–724.

Sutton, J., 1987. Bargaining Experiments, *European Economic Review*, 31, 272–284.

Tullock, G., 1999, Non-Prisoner's Dilemma, *Journal of Economic Behavior and Organization*, 39, 455–458.

Weg, E. and R. Zwick, 1999. Infinite Horizon Bargaining Games: Theory and Experiments. In D. Budescu, I. Erev, and R. Zwick, editors, *Games and Human Behavior: Essays in Honor of Amon Rapoport*, Laurence Erlbaum Associates, Mahwah, NJ.

Wua, J-J, Zhang, B-Y, Zhou, Z-X, He, Q-Q, Zheng, X-D, Cressman, R. and Y. Tao, 2009. Costly Punishment does not Always Increase Cooperation, *Proceedings of the National Academy of Science*, 106, 17448–17451.

Yang, C-L, Yue, C-S J. and I-T Yu, 2007. The rise of cooperation in correlated matching Prisoner's Dilemma: An experiment, *Experimental Economics*, 10, 3–20.

Appendix 1: Pre-play strategies as tacit messages – Treatment GCP-*Asym*

Dialogue 1 (*S*4):
Common beliefs on cooperativeness

Proposer: Would you be nice to me if I were nice to you?
Responder: Yes!
Proposer: Then, let's cooperate.

No. of pairs	Period	Proposer	Responder	Confirmation
17	1	**B**	**B**	**Yes**

Dialogue 2 (*S*2-*Nash*):
P tests R's strategic competence but P is (extremely) impatient.

Proposer: Are you smart?
Responder: Yes, I am!
Proposer: I am not patient.

No. of pairs	Period	Proposer	Responder	Confirmation
1	1	**A**	**A**	**Yes**

Dialogue 3 (*S*1):
P tests R's strategic ability and R is strategically incompetent or extremely impatient.

Proposer: Are you smart?
Responder: No, I am not!
Proposer: Then, I'll take advantage of this!

No. of pairs	Period	Proposer	Responder	Confirmation
2	1	**A**	**B**	**Yes**

Dialogue 3-bis (*S*1 & *S*4):
P tests R's strategic ability and R is strategically incompetent or extremely impatient; but P does not take advantage of this.

Proposer: Are you smart?
Responder: No, I am not!
Proposer: I don't want to take advantage of this!
Proposer: Would you be nice to me if I were nice to you?
Responder: Yes!
Proposer: Then, let's cooperate.

No. of pairs: 1

Period	Proposer	Responder	Confirmation
1	A	B	No
2	**B**	**B**	**Yes**

Dialogue 4 (*S*2 & *S*4):
P tests R's strategic competence and R is strategically competent.

Proposer: Are you smart?
Responder: Yes, I am!
Proposer: Would you be nice to me if I were nice to you?
Responder: Yes!
Proposer: Then, let's cooperate.

No. of pairs: 9

Period	Proposer	Responder	Confirmation
1	A	A	No
2	**B**	**B**	**Yes**

No. of pairs: 2

Period	Proposer	Responder	Confirmation
1-2	A	A	No
3	**B**	**B**	**Yes**

No. of pairs: 1

Period	Proposer	Responder	Confirmation
1-22	A	A	No
23	**B**	**B**	**Yes**

No. of pairs: 1

Period	Proposer	Responder	Confirmation
1-60	A	A	No
61	**B**	**B**	**Yes**

Dialogue 5 (*S*3-*A* & *S*4):
R tests P's strategic competence and P is strategically competent.

Proposer: Would you be nice to me if I were nice to you?
Responder: Wait a minute! Are you smart?
Proposer: I am smart!
Proposer: Would you be nice to me if I were nice to you?
Responder: Yes!
Proposer: Then, let's cooperate.

No. of pairs: 1

Period	Proposer	Responder	Confirmation
1	B	A	No
2	**B**	**B**	**Yes**

Dialogue 6 (*S*2, *S*3-A & *S*4) :
P tests R's strategic competence and R tests P's strategic competence;
both P and R are strategically competent.

No. of pairs: 2

Period	Proposer	Responder	Confirmation
1	A	A	No
2	B	A	No
3	**B**	**B**	**Yes**

No. of pairs: 1

Period	Proposer	Responder	Confirmation
1	A	A	No
2-3	B	A	No
4	**B**	**B**	**Yes**

No. of pairs: 1

Period	Proposer	Responder	Confirmation
1	A	A	No
2-4	B	A	No
5	**B**	**B**	**Yes**

No. of pairs: 1

Period	Proposer	Responder	Confirmation
1	B	A	No
2	A	A	No
3	**B**	**B**	**Yes**

No. of pairs

1

Period	Proposer	Responder	Confirmation
1	A	A	No
2	B	A	No
3	A	A	No
4	**B**	**B**	**Yes**

No. of pairs

1

Period	Proposer	Responder	Confirmation
1	A	A	No
2-6	B	A	No
7	A	A	No
8	**B**	**B**	**Yes**

No. of pairs

1

Period	Proposer	Responder	Confirmation
1-11	A	A	No
12-13	B	A	No
14	A	A	No
15	B	A	No
16	A	A	No
17	B	A	No
18-20	A	A	No
21	B	A	No
22-47	A	A	No
48	B	A	No
49	A	A	No
50	B	A	No
51-53	A	A	No
54-60	B	A	No
61-64	A	A	No
65	**B**	**B**	**Yes**

Dialogue 6-bis (*S*2, *S*3-*A* & *S*2-*Nash*):
P tests R's strategic competence and R tests P's strategic competence;
P (finally) punishes R's extreme toughness.

No. of pairs

1

Period	Proposer	Responder	Confirmation
1	A	A	No
2-5	B	A	No
6	A	A	No

7-10	B	A	No
11-12	A	A	No
13	B	A	No
14-15	A	A	No
16-17	B	A	No
18-19	A	A	No
20-23	B	A	No
24	A	A	No
25-29	B	A	No
30	A	A	No
31-32	B	A	No
33	**A**	**A**	**Yes**

Dialogue 7 (*S*2, *S*3-*B* & *S*4):
P tests R's strategic competence and R's cooperativeness.

Proposer: Would you be nice if I were nice to you?
Responder: Yes!
Proposer: Wait a minute! Are you clever?
Responder: I am clever!
Proposer: Would you be nice if I were nice to you?
Responder: Yes!
Proposer: Then, let's cooperate.

No. of pairs
1

Period	Proposer	Responder	Confirmation
1	B	B	No
2	A	A	No
3	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1	A	A	No
2	B	B	No
3	A	A	No
4	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1	A	A	No
2	B	B	No
3-5	A	A	No

Period	Proposer	Responder	Confirmation
6	B	B	No
7-8	A	A	No
9	B	B	No
10-11	A	A	No
12	B	B	No
13	A	A	No
14	B	B	No
15-16	A	A	No
17	B	B	No
18-24	A	A	No
25	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1-2	A	A	No
3	B	B	No
4-11	A	A	No
12	B	B	No
13-29	A	A	No
30	**B**	**B**	**Yes**

Dialogue 8:
'Trembling-hand Temptation' (cheap talk):
P tests R's strategic competence and R tests P's strategic competence;
both P and R are strategic competent and (finally) cooperates.

No. of pairs
1

Period	Proposer	Responder	Confirmation
1	A	A	No
2	B	B	No
3-4	A	A	No
5	B	B	No
6	A	A	No
7	B	B	No
8	A	A	No
9	B	A	No
10	A	A	No
11	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1	A	A	No

2	B	A	No
3-4	B	B	No
5	A	A	No
6	B	B	No
7-8	A	A	No
9	B	A	No
10	A	A	No
11	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1-3	A	A	No
4	B	B	No
5-7	A	A	No
8-12	B	A	No
13-15	A	A	No
16	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1-6	A	A	No
7	B	B	No
8	B	A	No
9	B	B	No
10-23	A	A	No
24	B	A	No
25	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1-7	A	A	No
8	B	A	No
9-29	A	A	No
30	B	B	No
31-37	A	A	No
38	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1-6	A	A	No
7	B	A	No
8-16	A	A	No
17-18	B	B	No

	19-37	A	A	No
	38	B	B	No
	39-40	A	A	No
	41-44	B	B	No
	45-47	A	A	No
	48	B	B	No
	49-55	A	A	No
	56	B	A	No
	57-58	A	A	No
	59	B	B	No
	60-62	A	A	No
	63	B	A	No
	64	**B**	**B**	**Yes**

Appendix 2: Pre-play strategies as tacit messages – Treatment GCP-*Sym*

Dialogue 1 (*S*4):
Common beliefs on cooperativeness

Question: Would you be nice to me if I am nice to you?
Answer: Yes!
Agreement: Then, let's cooperate.

No. of pairs: 6

Period	Proposer	Responder	Confirmation
1	**B**	**B**	**Yes**

Dialogue 2 (*S*2-*Nash*):
(At least one) P tests R's strategic competence and (the last) P is extremely impatient.

… … …
Question: Are you smart?
Answer: Yes, I am!
Agreement: I am not patient.

No. of pairs: 1

Period	Proposer	Responder	Confirmation
1	**A**	**A**	**Yes**

No. of pairs: 1

Period	Proposer	Responder	Confirmation
1	A	A	No
2	**A**	**A**	**Yes**

Dialogue 3 (*S*1):

(At least one) P tests R's strategic competence and (the last) R is strategically incompetent or extremely impatient.

… … …

Question: Are you smart?

Answer: No, I am not!

Agreement: Then, I'll take advantage of this!

No. of pairs: 1

Period	Proposer	Responder	Confirmation
1	**A**	**B**	**Yes**

No. of pairs: 1

Period	Proposer	Responder	Confirmation
1	A	A	No
2	**A**	**B**	**Yes**

Dialogue 4 (*S*2 & *S*4):

(At least one) P tests R's strategic competence and R is strategically competent.

… … …

Question: Are you smart?

Answer/Question: Yes, I am!

Question: Would you be nice to me if I were nice to you?

Answer: Yes!

Agreement: Then, let's cooperate.

No. of pairs: 8

Period	Proposer	Responder	Confirmation
1	A	A	No
2	**B**	**B**	**Yes**

No. of pairs: 5

Period	Proposer	Responder	Confirmation
1-2	A	A	No
3	**B**	**B**	**Yes**

No. of pairs: 1

Period	Proposer	Responder	Confirmation
1-3	A	A	No
4	**B**	**B**	**Yes**

Dialogue 5 (*S*3-*A* & *S*4):
(At least one) R tests P's strategic competence and P is strategically competent.

Question: Would you be nice to me if I were nice to you?
Answer/Question: Wait a minute! Are you smart?
Answer: I am smart!
Question: Would you be nice to me if I were nice to you?
Answer: Yes!
Agreement: Then, let's cooperate.

No. of pairs
3

Period	Proposer	Responder	Confirmation
1	B	A	No
2	**B**	**B**	**Yes**

Dialogue 5-bis (*S*3-*A*):
(At least one) R tests P's strategic competence and (the last) P is strategically incompetent.

… … …
Question: Would you be nice to me if I were nice to you?
Answer/Question: Wait a minute! Are you smart?
Answer/Agreement: No, I am not. You take advantage of this!

No. of pairs
1

Period	Proposer	Responder	Confirmation
1	B	A	No
2	**B**	**A**	**Yes**

Dialogue 6 (*S*2, *S*3-A & *S*4):
(At least one) P tests R's strategic competence and (at least one) R tests P's strategic competence; both P and R are strategically competent.

No. of pairs
4

Period	Proposer	Responder	Confirmation
1	A	A	No
2	B	A	No
3	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1-2	A	A	No
3	B	A	No
4	**B**	**B**	**Yes**

No. of pairs
6

Period	Proposer	Responder	Confirmation
1	B	A	No
2	A	A	No
3	**B**	**B**	**Yes**

No. of pairs
3

Period	Proposer	Responder	Confirmation
1	A	A	No
2	B	A	No
3	A	A	No
4	**B**	**B**	**Yes**

No of pairs
1

Period	Proposer	Responder	Confirmation
1-2	A	A	No
3	B	A	No
4-5	A	A	No
6	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1-2	A	A	No
3	B	A	No
4	A	A	No
5	B	A	No
6	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1-3	A	A	No
4	B	A	No
5	A	A	No
6	B	A	No
7	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1	B	A	No
2-4	A	A	No
5	B	A	No
6	A	A	No
7	B	A	No
8	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1	A	A	No
2	B	A	No
3	A	A	No
4	B	A	No
5	A	A	No
6	B	A	No
7	A	A	No
8	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1	A	A	No
2	B	A	No
3	A	A	No
4	B	A	No
5	A	A	No
6	B	A	No
7-9	A	A	No
10	**B**	**B**	**Yes**

Dialogue 8:
'Trembling-hand Temptation' (cheap talk):
(At least one) P tests R's strategic competence and (at least one) R tests P's strategic competence; both P and R are strategic competent and (finally) cooperates.

No. of pairs
2

Period	Proposer	Responder	Confirmation
1	A	A	No
2	B	B	No

3	B	A	No
4	A	A	No
5	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1	B	A	No
2	B	B	No
3	B	A	No
4	A	A	No
5	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1-2	A	A	No
3	B	A	No
4	B	B	No
5	B	A	No
6	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1	A	A	No
2	B	B	No
3	B	A	No
4	A	A	No
5	B	A	No
6	**B**	**B**	**Yes**

No. of pairs
1

Period	Proposer	Responder	Confirmation
1	B	A	No
2	A	A	No
3	B	A	No
4	B	B	No
5	B	A	No
6	**B**	**B**	**Yes**

1 \ 2	*A*	*B*
A	(**10**, **10**)	(25, 5)
B	(5, 25)	(15, 15)

Figure 1. Payoff matrix of the one-shot Prisoner's Dilemma

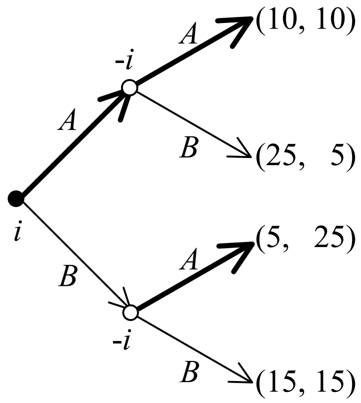

Figure 2. Two-stage Prisoner's Dilemma

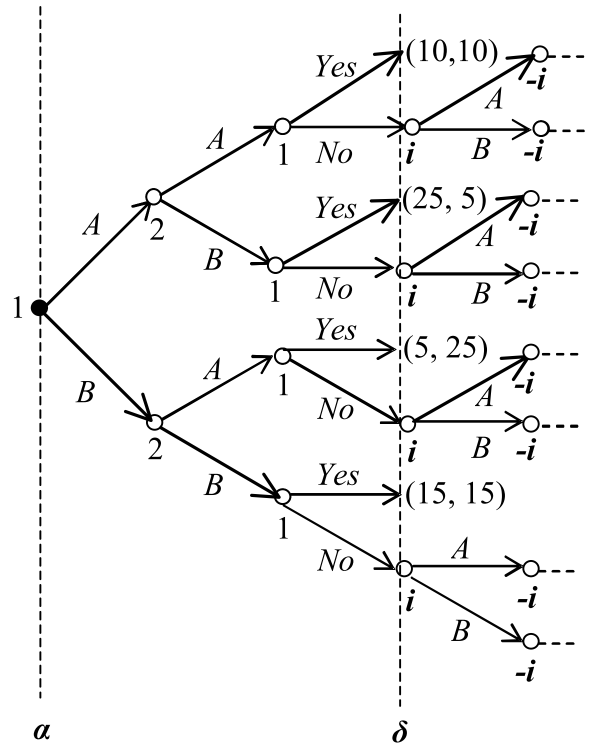

Figure 3. Prisoner's Dilemma with confirmed proposals

Outcome	No. of pairs in GCP-*Sym*	No. of pairs in GCP-*Asym*
Cooperation: (B, B)	49	50
Nash: (A, A)	2	2
Proposer ‘grabs’: (A, B)	2	2
Responder ‘grabs’: (B, A)	1	0
TOTAL	54	54

Table 1. Outcomes and confirmed agreements in treatments GCP-*Sym* and GCP-*Asym*.

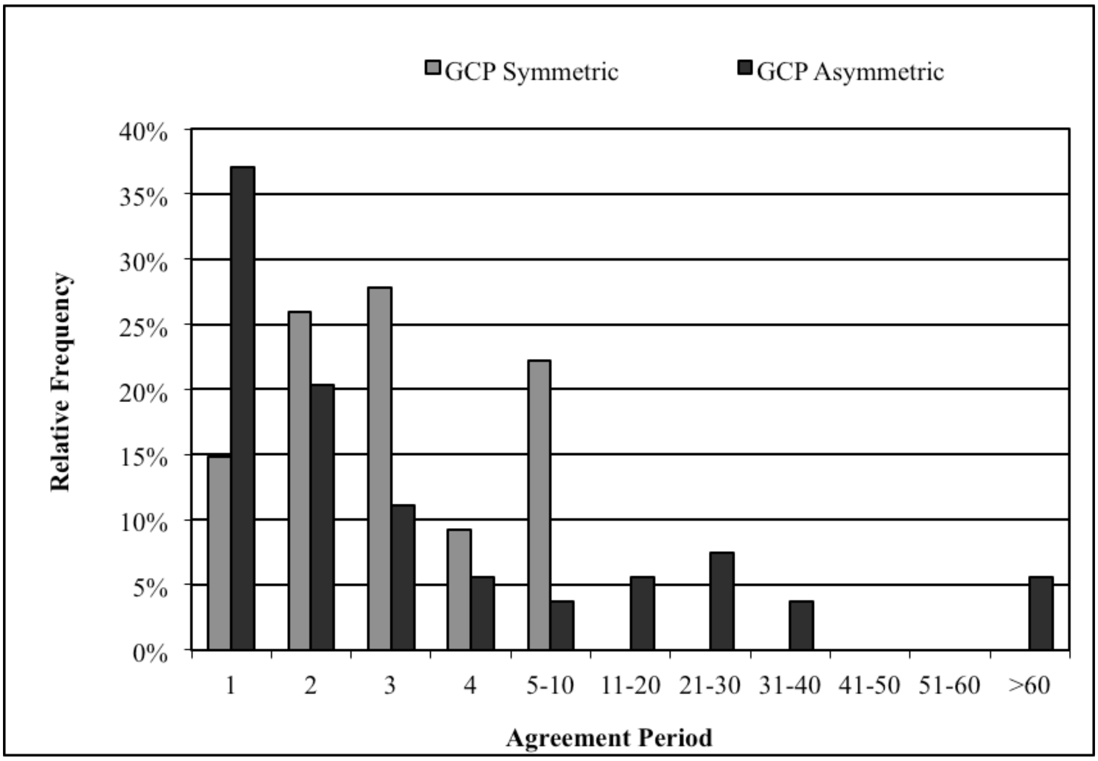

Figure 4. Distribution of the agreement period for GCP-*Sym* and GCP-*Asym*.