Marco Faillo, Alessandra Smerilli and Robert Sugden

The roles of level-k and team reasoning in solving coordination games

CEEL Working Paper 6-13

Cognitive and Experimental Economics Laboratory

Via Inama, 5 38100 Trento, Italy

http://www-ceel.economia.unitn.it
tel. +39.461.282313

The roles of level-k and team reasoning in solving coordination games.

Marco Faillo[a] , Alessandra Smerilli[b], Robert Sugden[b]

Abstract *Level-k and team reasoning theories, among others, have been used to explain experimental evidence on coordination games. Both theories succeed in explaining some results and both fail in explaining other results. Sometimes it is impossible to discriminate between them. For this reason we propose an experiment with pie games, similar to the ones used by Crawford et al. (2008). We observe subjects playing a series of coordination games, with different configurations of equality and Pareto-dominance, for which it is possible to provide clear predictions derived from both team reasoning and a particular cognitive hierarchy model: level-k theory. In line with previous experimental results, we find that each theory fails to predict observed behaviour in some games. However, because of the design of our experiment, we can go deeper into the matter. Our results show that Pareto dominance, fairness and uniqueness are good predictors for coordination choices. Secondly, we find mixed evidence about level-k and team reasoning theories. In particular team reasoning theory fails to predict choices when they picks out a solution which is Pareto dominated and not compensated by grater equality; Level-k theory fails in games in which it predicts the choice of one of not unique slices, and the unique choice is more equal than the alternative choices. This could represent a step forward to investigate the presence of team reasoning or level-k in coordinating behaviour.*

Keywords: Coordination games, Focal points, Team reasoning, Level-k theory.

JEL Classification: C72, C91, A13.

[a]Department of Economics and Management, University of Trento.

[b]School of Economics, University of East Anglia.

1. Introduction

That people can coordinate their actions in one-shot games with several Nash equilibria is no more a mystery in game theory. What is still under investigation is *how* they coordinate. Experimental evidence from pure coordination, Hi-Lo and battle of the sexes games shows that players often coordinate successfully, although the coordination rate depends on some features of the games (see Camerer 2003 for a review).

Several explanations of coordination in equilibrium selection have been put forward in the recent literature. Among these, two main approaches are emerging: team reasoning and cognitive hierarchy theories.

According to team reasoning, players look for the equilibrium that is best for the players as a 'team'[1].

In very general terms, in cognitive hierarchy models players aim at maximizing their payoff and their reasoning is grounded on beliefs about what opponents of lower cognitive level would do.

[1] Colman et al. (2008) explains how team reasoning provides a justification for choosing payoff-dominant equilibria, a concept introduced by Harsanyi and Selten (1998).

Players are assumed to be heterogeneous in terms of cognitive levels. Thus naïve level 0 players will choose at random. Level 1 players will best respond to expected level 0's choice, level 2 player will best respond to expected level 1's choice and so on. The experimental evidence on the emergence of focal points in simple coordination games (see Metha et al. 1994, Bardsley et al. 2010, Crawford et al. 2008 and Isoni et al. 2012) reports mixed results about the relative merits of these two explanations: some results can be explained by both theories, some only by team reasoning, and some others only by cognitive hierarchy models.

Despite the inconclusive findings, it is possible to infer some clues from the literature. It appears that there are some characteristics of the equilibria in these games, independent of the two theories, which attract players, so that team reasoning or cognitive hierarchy predictions work better when they pick out attractive equilibria. Such characteristics may include Pareto dominance and equality of payoffs.

However, a formal test of this conjecture has not been provided yet.

This paper is an attempt to contribute to the solution of the puzzle. We observe subjects playing a series of coordination games, with different configurations of equality and Pareto-dominance, for which it is possible to provide clear predictions derived from both team reasoning and a particular cognitive hierarchy model: level-k theory. In line with previous experimental results, we find that each theory fails to predict observed behaviour in some games.

However, because of the design of our experiment, we can go deeper into the matter. In particular, we observe that team reasoning theory fails to predict choices when it picks out a solution which is Pareto dominated and not compensated by greater equality; level-k theory fails in games in which it predicts a choice which is less equal than the alternative choices.

Two alternative explanations can account for this evidence. One is related to Bacharach's (2006) theory of team reasoning: according to this explanation, team reasoning and individual reasoning are two modes of reasoning which can be activated by different characteristics of the games.

The other explanation is based on the assumption that players are team reasoners, but not every one is so sophisticated in his reasoning to follow all the steps team reasoning requires to reach a solution.

We call these players, who are mostly guided by Pareto dominance considerations, 'naïve' team reasoners. We show that allowing the presence of naïve team reasoners organizes our results very well.

Team reasoning and cognitive hierarchies theories, and experimental studies aimed at testing them will be analysed in section 2. In section 3 we present the experimental design and procedures, in section 4 we discuss the theoretical predictions, in section 5 we present and discuss the results and section 6 shows our conclusions.

2. Team reasoning and Level-k: experimental evidence and theoretical issues

2.1. Team reasoning and level-k theories

Team reasoning and cognitive hierarchy theories, as explanations of selection of equilibrium in coordination games, appeared as alternative explanation of focal points[2]. Mehta et al. (1994) distinguish between two main explanatory strategies. In one approach, which Mehta et al. attribute to Lewis (1969), players' choices are grounded on *primary salience* (i.e. psychological propensities to pick particular strategies by default) and *secondary salience* (i.e. players' beliefs about other players' perceptions of primary salience). In the other approach, attributed to Schelling (1960), players look for a "rule of selection (and by extension, the label or strategy that it identifies)… [which] suggests itself or seems obvious or natural to people who are looking for ways of solving coordination problems" (p. 661). A rule of this kind has *Schelling salience*. The first approach focuses on individual strategic reasoning, assuming that players, who differ in their cognitive abilities, aim at maximizing their payoffs by best replying to the strategy that they expect their opponents to play. The second approach assumes that the shared objective of the players is to reach coordination, and in order to do so they try to find an effective common rule of conduct.

Metha et al. report an experimental investigation of pure coordination games. Most of the findings of this experiment are compatible with both secondary salience and Schelling salience. They conclude: "Our results suggest that Schelling salience may be playing a significant role. A major priority must now be to construct a more formal theory of Schelling salience which will generate specific hypotheses that can be tested experimentally" (p. 682).

As applied to coordination games, team reasoning can be thought of as an attempt to provide this formal theory of Schelling salience, whereas cognitive hierarchy theory is a development of primary and secondary salience. However, both theories are more general than this.

Different general formulations of team reasoning (or 'we-reasoning') have been proposed by David Hodgson (1967), Donald Regan (1980), Margaret Gilbert (1989), Susan Hurley (1989), Raimo Tuomela (1995, 2007), and Martin Hollis (1998). Within this body of literature, Robert Sugden (1993, 2000, 2003) and Michael Bacharach (1995, 1997, 1999, 2006) have developed game-theoretic analyses.

[2] Focal points have been introduced by Schelling (1960): they are particular Nash Equilibria on which players' expectations converge.

The key idea is summarised by Bacharach as: "Roughly, somebody 'team-reasons' if she works out the best feasible combination of actions for all the members of her team, then does her part in it" (Bacharach 2006, p. 121).

In other words, when people team-reason they seek an answer to the question: "What should we do?", and they act accordingly.

A shared view among scholars who study team reasoning is that in some circumstances people team reason, in some others not. Circumspect Team reasoning (Bacharach 2006), common reason to believe (Sugden 2003), game harmony (Tan and Zizzo 2008) and vacillation (Smerilli 2012) are models which try to explain this fact. However, why and when people team reason and why and when they do not remains still unclear.

Level-k and cognitive hierarchy theories (Stahl and Wilson (1994, 1995); Nagel (1995); Ho, Camerer and Weigelt (1998); Bacharach and Stahl (2000); Costa-Gomes, Crawford and Broseta (2001); Camerer, Ho and Chong; Costa-Gomes and Crawford (2006)) can be thought as formalized models of strategic reasoning based on primary and secondary salience.

In this work we concentrate on Level-k theory[3]. In these models, each player belongs to a category (type) and follows a rule. In general a type L1 will anchor his/her beliefs in a nonstrategic L0 type, and best respond to this. A L2 player best responds to L1, and so on. Then k, k=1,2,3,... captures the level of reasoning. Thus the behaviour of players at all levels above L0 is grounded in beliefs about L0 behaviour. The behaviour of the nonstrategic type L0 is different in different versions of the theory. In some versions, L0 chooses at random, which implies that L1 takes account only of his own payoffs; in other versions, L0 follows 'payoff salience', which means that he takes account only of his own payoffs; in still other versions, L0 favours primarily salient' labels.

2.2. Experimental evidence

Experimental evidence on coordination games shows mixed results: sometimes it seems that subjects act according to team reasoning theories, sometimes according to level-k.

Crawford et al. (2008) report experiments using pure coordination games with labels, battle of the sexes (with and without labels) and "pie" games. In a pie game, subjects try to coordinate by choosing the same 'slice' of a three-slice pie. Different slices have different payoff combinations, and one slice is coloured differently from the other two. Crawford et al. propose a level-k model that explains the evidence from many of these games, but note that the choices made in some pie games can be explained only by team reasoning.

---

[3] Cognitive Hierarchy Models and level-k theory, applied to our games, give the same qualitative predictions.

Bardsley et. al (2010) report experimental evidence about behaviour in pure coordination games and Hi-Lo games. These experiments are run, with apparently minor variations, in two different places: the results from Amsterdam seem to support team reasoning, whereas the results from Nottingham can be explained by level-k theory.

Isoni et al. (2012) investigate games with the same payoff structures as those of Crawford et al's. pure coordination and battle of the sexes games, but with different displays. They too find mixed (but less extreme) results.

By going deeper into this literature, at least three clues can be inferred.

The first clue is that *team reasoning predictions may be less likely to work when two or more equilibria are not Pareto ranked and team reasoning predicts the choice of one of these*. In such games, team reasoning has to deal with a conflict of interest between players.

Crawford et al. (2008) compare pure coordination games with battle of the sexes games, when both are presented with the same labelling. They find high rates of coordination in pure coordination games, but coordination fails in battle of sexes.

A similar but less strong result is obtained by Isoni et al.(2012), who find that although focal points work in battle of sexes, they are less effective than in pure coordination games. This suggests that there is more individualistic reasoning when there is a conflict of interests.

A second clue is that *equality can favour team reasoning*. If team reasoning recommends a solution with equal payoffs, this solution is liable to be chosen even if level-k recommends another solution.

Crawford et al. (2008: 1456) report two pie games in which the slice that is distinguished by colour has the payoffs (5, 5). In 'game AM1' the other two slices have payoffs (5, 6) and (6, 5); in 'game AL1', these payoffs are (5, 10) and (10, 5). Contrary to level k theory, but consistently with team reasoning, most subjects choose the (5, 5) slices in these games.

A third clue is that when there is a *conflict between ex-post Pareto-dominance and ex-ante Pareto-dominance, team reasoning predictions* of ex-ante Pareto-dominant solutions *can fail to work.*

Consider for example one of the 'number task' games proposed by Bardsley et al. (2010) in which two subjects must coordinate by choosing the same option among:

(10,10) (10,10) (10,10) (9,9)[4].

Team reasoning recommends (9,9), because is 'unique'. If the players cannot distinguish between the (10,10) options, there is no rule which can guarantee that their payoffs will be (10,10). So, from

[4] The payoffs are not displayed in a line, but they have a neutral display.

an ex ante perspective, and provided that the players are not extremely risk-loving, the rule "choose (9,9)" Pareto dominates the rule "pick a (10,10)". Then, if subjects ask themselves 'what should we do?', it is evident that (9,9) is the best choice for "us".

Ex-post, however, once the choice is made and the other's choice is known, (9,9) is Pareto dominated by (10,10). When this task was used in Bardsley et al.'s Amsterdam experiment, most subjects coordinated on (9,9), in a similar Nottingham experiment, most subjects distributed their choices over the (10,10) options, as predicted by level k theory.

These clues have been noticed already.

Crawford et al. (2008) allude to the first two clues when they conclude their paper with a conjecture: "We speculate that the use of team reasoning depends on Pareto-dominance relations among coordination outcomes and their degree of payoff conflict"(p. 1456). With regard to the third clue, Bardsley et al. speculate that there was some tendency for the modes of reasoning used in previous pure coordination games (different in Amsterdam and Nottingham) to spill over to the number tasks. Because the focal points in the Amsterdam pure coordination games were 'odd ones out', this may have primed players to think of the unattractive uniqueness of the (9, 9) option as a means of coordination. The suggestion seems to be that the possibility of using ex ante Pareto dominance as a coordination device is not immediately obvious to many subjects.

Although we can infer these conjectures from the literature, the results on which they are based are not systematic: for this reason we carry out a controlled test in which every game has a unique team reasoning choice and the relationship between Pareto dominance and equality varies between games. We use unlabelled games, in order to produce a more controlled test, by reducing the number of potential explanatory variables.

3. The experiment: design, theoretical predictions and procedures

As has been seen in the previous section, it is not always clear how and when team reasoning and level-k theories work as explanations of coordinating behaviour. The experiment is aimed at discriminating between the two theories as explanation of coordination in simple games. Moreover, it allows to investigate the three clues discussed in the previous section.

For this reason, the experiment focuses on two relevant characteristics of equilibria: equality and Pareto dominance. By using games with different configurations of Pareto dominance and equality, we are able to investigate the relative power of these characteristics to attract players to particular equilibria.

3.1 The pie games

The experiment uses two-person pie games similar to those used by Crawford et al (2008), except that all the slices have the same colour. Payoffs are chosen so that the predictions of both team reasoning and level-k theory are unambiguous, with a unique team reasoning optimal choice in each game.

Formally, each game is a 3x3 coordination game with the payoff matrix shown in Figure 1. The parameters x, y, v, w are always strictly positive and satisfy $y \geq x$ and $\{v, w\} \neq \{x, y\}$. The last condition ensures that the strategy pair (R3, R3) is *unique* in the sense that it can be distinguished from all other pairs by reference only to payoffs. In contrast, (R1, R1) and (R2, R2) are symmetrical and so non-unique.

[FIGURE 1]

Figure 2 shows how a typical game was seen by subjects. (This is a game with x=9, y=10, v=9, w=9.) Three different displays of this game are shown, corresponding to three different treatments A,B,C . The labels 'R1', 'R2' and 'R3' were *not* seen by subjects. In each treatment, both co-players see the same pie divided into three slices. Each player independently chooses one of the slices. If their choices coincide, they get the payoffs that appear in the slice; otherwise they both get nothing. Notice that, because players are referred to as 'you' and 'the other', there is no commonly known and payoff-independent labelling that distinguishes between the players.

Because of this, the only labelling feature that distinguishes R1 and R2 from one another is the positions of their slices in the pie.

[FIGURE 2]

Our working assumption is that slice positions are *nondescript* in the sense of Bacharach (2006, p. 16). That is, descriptions (such as 'left slice') that in principle could be used to pick out particular equilibria do not easily come to mind to normal players. Thus, players cannot solve the problem of coordinating on one of the slices R1 and R2 rather than the other. By comparing behaviour in the three treatments, which differ only in the positioning of the slices, we will be able to test this assumption.

The experiment investigated eleven games, G1 to G11. The payoffs that define these games are shown in Table 1.

[TABLE 1]

These games have different configurations of Pareto dominance and equality. Games in which R1 and R2 give equal payoffs (i.e. x=y) are shown by 'E' in the 'R1/R2' column. Games in which R1 and R2 weakly Pareto dominate R3 (i.e. x,y ≥ v and x,y ≥ w with at least one strict inequality) are shown by 'P' in this column. Games in which R3 gives equal payoffs (i.e. v=w) are shown by 'E' in the 'R3' column. Games in which R3 weakly Pareto dominates R1 and R2 (i.e. v ≥ x,y and w ≥ x,y with at least one strict inequality) are shown by 'P' in this column. By using 11 pie games we include all the possible combinations of entries (i.e. 'E', 'P', 'E,P' and '-') in the two columns[5].

3.2 *Theoretical predictions*

The experiment is designed to test if and when players act according to team reasoning or level-k theories. In this section we analyse the theoretical predictions for each game, using the assumption that slice positions are nondescript.

3.2.1 Team reasoning predictions

In our pie games R3 differs from R1 and R2, because R1 and R2 are always symmetrical. What is the team optimal choice in this case? We shall prove that R3 is the team optimal choice in each game, but firstly we give an informal intuition for this result.

Because R3 is unique, (v,w) is a payoff combination that the players, acting as a team, are able to obtain with certainty. But if the players have no commonly-understood means of coordinating on one of R1 and R2, the only rule they can use is 'pick one of R1 and R2', which represents a lottery for the team. Given the payoffs in our games, any team that was not extremely risk-loving would choose R3 rather than this lottery, even if R3 was Pareto-dominated by each of R1 and R2 separately.

To formalise this intuitive argument, we begin by defining a group utility function U, following Bacharach (2006, pp. 87-88). In general, one of the problems in defining such a function is to deal with inequality and risk aversion, but the parameters used in our games make the implications of team reasoning insensitive to these characteristics.

Following the literature on team reasoning, we make the following three assumptions about U.

[5] Note that because of the condition {v, w} ≠ {x, y}, needed to ensure R3 is unique, we cannot have 'E' (as distinct from 'E, P') in both columns. Nor, because of the definition of Pareto dominance, can we have 'P' or 'E,P' in both columns.

First, we assume that U is *symmetrical*, that is, for all payoffs s, t, U(s,t) = U(t,s). According to Bacharach (2006), "It is reasonable to suppose that principles of symmetry between individual payoffs will be respected in U" (p.145). This property means that, when engaging in team reasoning, each player treats his own payoffs in exactly the same way as his co-player's.

Secondly, we assume *increasingness*, that is, U(s,t) is increasing in s and t. This assumption is used by Bacharach, who calls it the 'Paretianness Condition'.

Finally, we assume that the group utility function has the property of risk aversion (that is, is concave in its arguments). Risk aversion requires that: U(s,t)>U(2s,2t)/2.

This seems a natural assumption, although Bacharach did not mention it in his work. It is used by Bardsley et al (2010)[6].

We now show that these assumptions imply that, if the players are unable to coordinate on one of R1 and R2, then R3 is the team optimal option.

If we normalise U(0,0) = 0, then (using symmetry), R3 is team optimal if U(v,w)>U(x,y)/2. Without loss of generality, let v≥w and y≤x. Then a sufficient condition for R3 to be team optimal is U(w,w)>U(y/2,y/2). By risk aversion, we have that U(y/2,y/2)>U(y,y)/2. So a sufficient condition for R3 to be team optimal is U(w,w)> U(y/2,y/2). By increasingness and symmetry, this is equivalent to w>y/2 or w/y >1/2. In our games (see Table 1) the lowest value of w/y is 8/10. Thus R3 is (very strongly) team optimal in every game.

3.2.2 Level-k prediction

To make predictions of level-k individual reasoning, we need to make assumptions about L0 players. The usual assumption in level-k models is that L0 choices are random, but in the model of Crawford et al., L0 responds to payoff salience and label salience, with a bias for payoff salience. In our games there are no labels, so it is impossible to follow label salience. Initially, let us assume that L0 plays at random.

Thus, at L0 in every game, for both players, pr (R1) = pr(R2) = pr(R3)= 1/3. At each higher level, each player chooses a best reply to a co-player at the immediately lower level.

For example, consider game 1. At L0, players randomise over R1, R2 and R3. At L1, player 1 best replies to an L0 co-player by choosing R2, and similarly, player 2's best reply is R1. At L2, player 1's best reply to an L1 co-player is R2, and similarly, player 2's best reply is R1. At L3, player 1 chooses R1 and player 2 chooses R2; and so on (see Table 2).

[TABLE 2]

[6] See Gold 2012 for a review of Utility group function properties.

Generalising, at L1 each player makes a best reply to L0. This is as if randomising over strategies that are optimal for her under the assumption that the other player randomises. At L2 each player eliminates strategies not chosen by her co-player at L1, then optimises as at L1. So, if for any player there is a unique choice at any level, this repeats itself for her co-player at the next level up, etc.

This principle operates at L1 for the following games:

- in G2 and G6, both players choose R3 at L1 and this repeats itself at every higher level;
- in G1, G3, G9 and G11, one of the players has R1 as unique choice, the other has R2, and this repeats itself at every higher level.

Also, if at L1 both players randomise over R1 and R2, this repeats itself at every higher level. This happens in G5 and G7.

In G4 and G8 we observe convergence to R3 from L2.

G10 is different: at L1, one player has R3 as the unique choice and the other randomises over R1 and R2. This pattern repeats itself indefinitely. So at every level above L0, averaging over the two players, p(R3)= ½ and p(R1) = p(R2) = ¼ .

These predictions are based on the assumption that at L0 players choose at random. If, instead, we use the 'payoff salience' specification of level-k theory, L0 will choose the slice with the highest own payoff, i.e. L0 behaves in the same way as L1 does in the random specification; L1 behaves like L2 and so on. This means that in the two cases the predictions are very similar, but in the 'payoff salience' specification they are sharper, in the sense that convergence takes place at a lower level, as can be seen in table 3 for game 1.

[TABLE 3]

It is worth noticing that the proportion of R3 choices made by L1, and hence by all higher levels, is 0 if $y > v, w$ (remember that $y \geq x$). This happens in games G1, G3, G5, G7, G9 and G11. A sufficient condition for this result is that R1, R2 ex-post Pareto-dominate R3, i.e. $x \geq v, w$.

The proportion of R3 choices made by L1 is 1 if $v, w > y$. This happens in G2 and G6. In addition, the proportion of R3 choices made by L2, and then by all higher levels, is 1 if $v, w \geq y$. This happens in G4 and G8. So if R3 ex post Pareto-dominates R1, R2, level-k predicts a high frequency of R3 choices.

Team reasoning and Level-k theory's predictions for games 1-11 are reported in table 4. The 'predicted proportion of R3 choices' averages over players 1 and 2. In every case, and independently of the distribution of levels, each theory *either* predicts that this proportion is strictly greater than 1/3 *or* predicts that this proportion is strictly less than 1/3. Since completely random choice would produce a 1/3 proportion, these predictions would not be affected by adding noise to the model.

[TABLE 4]

3.3 *Hypotheses*

Team reasoning and level-k theoretical predictions differ in games G1, G3, G5, G7, G9 and G11, in which team reasoning predicts a high proportion of R3 choices whereas level-k predicts that this proportion will be low. It follows that in these games it is possible to discriminate between the two theories. For each game we are able to test if the proportion of R3 choices is significantly higher or lower than 1/3. In the first case, the hypothesis of level-k reasoning can be rejected, whereas in the latter case, the hypothesis of team reasoning can be rejected.

Secondly, we are interested in the effect of Equality and Pareto dominance on team reasoning.

The experiment is designed to test the predictions of team reasoning under a range of different values of *v, w, x, y*. In particular we aim to test whether the tendency to choose the team optimal slice depends on ex post Pareto dominance and equality.

With regard to ex post Pareto dominance we can distinguish three cases:

a. (v,w) Pareto-dominates (x,y) and (y,x), i.e., R3 is team-optimal ex post as well as ex ante. This occurs in G2, G4, G6 and G8. This condition can be expected to favour R3.

b. (x,y) and (y,x) Pareto-dominate (v,w), i.e., R3 is Pareto-dominated ex post. This is the case in G1, G3, G5 and G7. This condition can be expected to disfavour R3.

c. (x, y) and (y, x) are not Pareto-ranked relative to (v, w), i.e. the case of 'conflict of interests'. This occurs in G9, G10 and G11. This condition can be expected to disfavour R3.

In relation to equality, the following cases can be distinguished:

d. v=w and x ≠ y, i.e., R3 is equal and R1 and R2 are unequal. This occurs in G1, G2 and G11. This condition can be expected to favour R3.

e. v ≠ w and x = y, i.e., R3 is unequal and R1 and R2 are equal. This occurs in G7, G8 and G10. This condition can be expected to disfavour R3.

f. v = w and x = y, i.e., R1, R2 and R3 are all equal. This occurs in G5 and G6. There seems no reason to expect this condition in itself either to favour or disfavour R3.

g. $v \neq w$ and $x \neq y$, i.e., R1, R2 and R3 are all unequal. This occurs in G3, G4 and G9. There seems no reason to expect this condition in itself either to favour or disfavour R3.

Case *c* corresponds to the first clue mentioned in section 2, case *d* corresponds to the second clue and case *b* to the third.

These conclusions are summarised in table 5.

[TABLE 5]

There are some games for which only one of the conditions *a* to *e* (i.e. conditions that favour or disfavour R3) holds. There are some for which two of these conditions hold, both working in the same direction (either favouring or disfavouring R3). And there are some for which two of these conditions hold, working in opposite directions.

In particular:

- In G3 and G5, the only relevant condition that holds is *b*. In G9 the only relevant condition that holds is *c*. In these games we would expect R3 to be disfavoured.
- In G4 and G6, the only relevant condition that holds is *a*. We would expect R3 to be favoured.
- In G2, the only relevant conditions that hold are *a*. and *d*. We would expect R3 to be favoured.
- In G7, the only relevant conditions that hold are *b*. and *e*. We would expect R3 to be disfavoured.
- In G10, the only relevant conditions that hold are *c*. and *e*. We would expect R3 to be disfavoured.
- In each of G1, G8 and G11, two relevant conditions hold (*b* and *d* in G1, *e* and *a* in G8, and *c* and *d* in G11), working in opposite directions.

The previous considerations represent a map, by the help of which we can look deeply into the evidence we obtain from the experiment.

3.4 Procedures

A total of 194 subjects participated voluntarily in the experiment at the CEEL Lab of the University of Trento. 10 sessions were conducted (8 with 20 participants, 1 with 18 participants and one with 16 participants) between June 2012 and November 2012. The experiment was programmed by using the z-Tree platform (Fischbacher, 2007). Subjects were undergraduate students (53.3% from economics and management, 42.5% females, 86.7% Italians). On their arrival at the laboratory, participants were welcomed and asked to draw lots, so that they were randomly assigned to terminals. Once all of them were seated, the instructions were handed to them in written form before being read aloud by the experimenter. The participants had to answer several control questions and we did not proceed with the actual experiment until all participants had answered all questions correctly.

Each subject played games G1 to G11, as described in Section 3.1, with payoffs expressed in Euros. Games were played anonymously. Co-players were re-matched between games. Different subjects played different sequences of games either as player 1 or player 2 . In particular, in each round each subject was assigned a game (from G1 to G11) in one of the three treatments A, B and C (see Appendix 1). For example, in a 20 subjects session, subject 1's first game was G1 in treatment 1, she played as player 1, with subject 6 as player 2.

Subject 2's first game was G11 in treatment B, she played as player 1with subject 7 as co-player; and so on.

No feedback was given until all eleven games had been played. At the end of the experiment one of the eleven rounds was randomly selected and subjects were paid according to the outcome of the game they played in that round. Subjects received also a show up fee of € 3. The average earning for each participant was € 6.50. Sessions averaged approximately 40 minutes.

4. Results

Table 6 reports the distribution of R1, R2 and R3 choices across treatments in the 11 games.

[TABLE 6]

First of all, remember that our team reasoning predictions depend on the working assumption that subjects cannot use position as a label to discriminate between R1, R2 and R3. In order to check this assumption we compare, for each game, the distributions of choices between R1, R2 and R3 in the three treatments (which differ only in the positions of the slices). Our test uses the null hypothesis that, for each game, the distribution of choices is the same across treatments (Pearson Chi-squared in Appendix 2).

In particular, our focus is on the proportions of R3 choices across treatments. Aggregating R1 and R2 choices we do not find any significant difference between treatments (see Appendix 2)[7].

We also test the null hypothesis that subjects chose at random. The null hypothesis is rejected in all the games except G10[8].

Now, with these premises, we can present the principal results of the experiment.

With regard to the theoretical predictions we can conclude that both theories fail in some games (see table 6).

Team reasoning fails in predicting R3 choice in G3, G5 and G7, where the actual proportion of R3 choices, averaged across players 1 and 2, is significantly smaller then 1/3.[9]

In all these game R3 is weakly or strictly Pareto dominated ex post: all these games satisfy condition *b* (see section 3.3) . G7 also satisfies condition *e* (R3 is unequal, whereas R1, R2 are equal). According to both conditions, R3 should be disfavoured.

However, condition *b* is also satisfied in G1 (R3 is weakly Pareto dominated), but in this game R3 is chosen by 74% of subjects. G1 differs from G3, G5 and G7 in that it satisfies both *b* (which disfavours R3) and *d* (which favours it). It seems as though the equality of R3 compensates for its being ex post Pareto dominated.[10]

Level-k fails in predicting that R3 will not be chosen in G1, G9, and G11, where the proportion of R3 choices, averaged across the two players, is significantly greater than 1/3. In G1 and G11, R3 is equal and R1 and R2 are not (i.e. condition *d* is satisfied). In G9, all three outcomes are unequal, but R3 is clearly *less* unequal than R1 and R2.

In the rest of the games both theories agree and they work well. The only exception is G10, in which both theories fail: the proportion of R3 choices is not significantly different from 1/3. This

---

[7] When we consider the disaggregated frequencies of R1, R2, and R3 choices, we find that the only systematic effect is that in treatment A, in games with x = y, R1 is chosen more frequently than R2. In some cases, test results are not reliable because the expected number of observations in some cells is too small, as in G2 and G6. The only case in which we have to reject the null hypothesis that the proportion of R3 choices is the same across treatments also by aggregating R1 and R2 is G9, in which x ≠y and the difference between treatments is the result of the behaviour of player 2 in treatment C.

[8] Chi squared test for goodness of fit:
Player 1 Treatment 1: Chi sq.=16.29, p = .0003; Player 1 Treatment 2: Chi sq.= 2.26, p = .3225; Player 1 Treatment 3: Chi sq.= 4.16, p = .1249;
Player 2 Treatment 1: Chi sq.= 7.82, p = .02; Player 2 Treatment 2: Chi sq.= 1.32, p = .5179; Player 2 Treatment 3: Chi sq.= 12.48, p = .2894

[9] In a two tail Binomial test with n = 140 and probability of success (choice=R3) = 0.33, 59 or more successes is significantly more than random and 35 or fewer is significantly less than random (5% level). With n = 194, the corresponding numbers are 79 and 53.

[10] This does not happen the other way round. In G8, R3 ex post Pareto dominates R1 and R2 (i.e., condition *a*, favouring R3) but is unequal, while R1 and R2 are equal (i.e., condition *e*, disfavouring R3). Here 52% of subjects choose R3.

game satisfies conditions *c* and *e*, both of which disfavour R3. Level-*k* theory implies that R3 is chosen with probability 1/2 at each level above L0.

5. Discussion

The results of the experiment agree with previous literature: in coordination games neither level-k theory nor team reasoning can explain behaviour in all the games. But differently from previous experiments, we obtain clearer and sharper results: we are able to identify general features of games in which team reasoning theory clearly fails in predicting choices, and of games in which it succeeds.

To sum up, team reasoning fails when it predicts the choice of a slice that is ex post Pareto dominated by the other two and this is not compensated by greater equality (games G3, G5 and G7). Team reasoning fails also (but not as badly) in G10, where the team-optimal choice is unequal and there is conflict of interests (condition *c*).

In order to explain the experimental evidence, we need a more general theory.

One possible line of explanation is based on Bacharach's theory. According to him, there exist two modes of reasoning: individual reasoning and team reasoning. So people sometimes team reason, sometimes not.

For Bacharach, modes of reasoning are not chosen rationally. The process by which a mode of reasoning comes into play is based on frames: if the we-frame comes to mind, the subject will group identify and then she will start to we-reason. A frame can be defined as a set of concepts that an agent uses when she is thinking about a decision problem. It cannot be chosen, and how it comes to mind is a psychological process:

> "Her frame stands to her thoughts as a set of axes does to a graph; it circumscribes the thoughts that are logically possible for her (not ever but at the time). In a decision problem, everything is up for framing... also up for framing are her coplayers, and herself" (ib. p. 69).

The we-frame, and therefore group identification, is favoured by certain characteristics of the games. This means that team reasoning can be activated when particular features are present. According to Bacharach (ib, pp. 82–83), one such feature is perceived interdependence, based on a recognition of 'common interest'. In a two-player game, the players have a *common interest* in some pair of strategies s* over some other pair s, if both prefer s* to s. Common interest, in this definition, is related to our concept of Pareto dominance. Another feature, called 'harmony of preferences' by Bacharach ib, pp. 63, 83), is related to the degree of conflict among payoffs. This idea has been developed by Zizzo and Tan (2008). They propose a measure of *game harmony*, based on correlation between payoffs, which is related to our concept of equality.

The results of our experiment seems to suggest that ex post Pareto dominance and equality play an important role in group identification; ex ante Pareto dominance, when conflicting with ex-post Pareto dominance, seems not to be sufficient.

Another possible explanation of our data is based on the assumption that team reasoners can be more or less sophisticated. When the R3 slice has both ex post Pareto-dominance and ex post equality, players can see that R3 is the best for the team even without using uniqueness. When, instead, ex-post Pareto-dominance and equality are not present, players are required to 'see' uniqueness; this means that they should be 'sophisticated' team reasoners.

So far we have considered the distinction between level-k reasoners and team reasoners. We assumed the latter to be sophisticated enough to recognize the uniqueness of R3 and to be aware of the distinction between ex-ante and ex-post Pareto dominance. However, we cannot exclude the existence of a different type of team reasoners, who, like the more sophisticated one, is willing to pursue the group interest, but at the same time does not recognize the uniqueness of R3 and adopts the simple rule of thumb of focusing on Pareto dominance and equality of the outcome. We call these agents 'naïve' team reasoners.

In Bacharach's circumspect team reasoning, there is space for individual and team reasoners. Team reasoners are aware of the presence of non team reasoners, and for this reason they maximize the utility of the team given the proportion of individual reasoners. In a similar way, to allow a presence of naïve team reasoners does not means that everybody is naïve: R1 and R2 choices could be the results of the presence of naïve team reasoners and of more sophisticated ones, who take in account the proportion of naïve reasoners.

Assuming that the team utility function is increasing and concave, naïve team reasoning organizes the data pretty well: it implies that R3 is the best option in G2, G4, G6, G9, G11 and R1 and R2 are the best options in G3, G5, G7, G10; the implication is not clear in G1 where R3 is equal but Pareto dominated by R1 and R2.

It is worth noticing that the existence of naïve team reasoners can also explain some experimental results reported in previous literature. Take for example game 'AM 4' (figure 3) reported by Crawford et al. (2008), which is very similar to game G3.

[FIGURE 3]

The sophisticated team reasoning solution for this game is B (for the same considerations we made in section 3.2.1). According to Crawford et al., level-*k* theory predicts L and B. In this game subjects chose L and R, which is exactly what a naïve team reasoner does, and this is also the way the authors explain the results. They say that 'B for both' is less equitable and weakly Pareto-dominated by 'R for both'. If we exclude B, the players are left with a 2x2 battle of sexes game in which they will alternate between R and L, because there is no way to break the symmetry.
Also, the choices of Nottingham subjects in the 'number task' games studied by Bardsley et al. (2010) can be explained by allowing the presence of naïve team reasoning. In these games there is a conflict between ex ante and ex post Pareto dominance. As we have already mentioned in section 2.2, according to the authors, in the Amsterdam version of the experiment most of the players coordinated on (9,9) because they were primed to see the uniqueness of choices. In Nottingham version, without any priming, it seems as though players behave like naïve team reasoners.
However, merely assuming the presence of naïve team reasoners would not explain coordination in pure coordination games and in games with conflict of interests (like battle of the sexes games). To solve a pure coordination games, team-reasoning players are required to re-describe the game in a way which requires some degree of sophistication. And experimental evidence shows that players are good at coordinating in pure coordination games.
At the same time, experimental evidence shows that it is more difficult to coordinate in battle of the sexes games than in pure coordination games. So if coordination is explained by team reasoning, we have to assume that team reasoning tends to be switched off by battle of the sexes games.
Overall, the evidence suggests that behaviour in coordination games might be explained by introducing a model containing both individual and team reasoners, with team reasoners having different levels of sophistication, activated by the characteristics of the games.

6. Conclusion

Our experiment was designed with two main objectives: to discriminate between level-k and team reasoning theories, and to investigate three clues, already present in previous literature, about the effects of Pareto ranking of payoffs, equality, and differences among ex ante and ex post Pareto dominance.
It represents a step forward to the understanding of coordinating behaviour. On the one hand, it confirms previous findings that neither team reasoning nor cognitive hierarchy models can completely explain experimental evidence. On the other hand it reveals how some characteristics of the equilibria in games, such as ex ante or ex post Pareto dominance and equality, can attract players.

As we have already mentioned in previous section, in order to explain the evidence, a more general theory is needed. Crawford et al. (2008) suggested that a 'judicious' combination of team reasoning and level-k theories, incorporating other considerations, is needed.

One contribution to this debate is to introduce naïve team reasoners in the team reasoning theory: our conjecture is that people can team reason in a more or less sophisticated way, depending on the characteristics of the games. Maybe the 'judicious' combination of team reasoning and level-k theories could result in the introduction of a sort of level-k team reasoning theory, in which both, individual and team reasoners are present and both with different levels of sophistication.

The answer is open: future works can verify this conjecture.

References

BACHARACH, M. (2006): Beyond Individual Choice. Princeton University Press, Edited by N. GOLD, and R. SUGDEN.

BACHARACH, M. (1999): "Interactive Team Reasoning: A Contribution to the Theory of Cooperation," Research in Economics, 53, 117-147.

BACHARACH, M. (1997): ""We" Equilibria: A Variable Frame Theory of Cooperation," Oxford: Institute of Economics and Statistics, University of Oxford, 30, available at http://cesswb.nuff.ox.ac.uk/documents/mb/Bacharach_WeEquilibria.pdf.

BACHARACH, M. (1995): "Co-Operating without Communicating," Working Paper, Institute of Economics and Statistics, University of Oxford.

BACHARACH, M. and STAHL, D.O. (2000): "Variable-frame level-n theory", Games and Economic Behavior, vol. 33(2), pp. 220–46.

BARDSLEY, N., MEHTA, J., STARMER, C., SUGDEN, R. (2010): "Explaining focal points: cognitive hierarchy theory versus team reasoning", Economic Journal 120 (March), 40–79.

CAMERER, C.F. (2003) *Behavioral Game Theory,* Princeton University Press.

CAMERER, C.F., HO, T.H. and CHONG, J.K. (2004): "A cognitive hierarchy model of games", Quarterly Journal of Economics, vol. 119(3), pp. 861–98.

COLMAN, A., PULFORD B. AND ROSE, J. (2008): "Collective rationality in interactive decisions: Evidence for team reasoning", Acta Psychologica, 128, 387-397.

COSTA-GOMES, M., CRAWFORD, V., (2006): "Cognition and Behavior in Two Person Guessing Games: An Experimental Study." American Economic Review, 96(5): 1737-
1768.

COSTA-GOMES, M., CRAWFORD, V., BROSETA, B., (2001): "Cognition and behavior in normal-form games: an experimental study", Econometrica 69, 1193–1235.

CRAWFORD, V.P., GNEEZY, U. and ROTTENSTREICH, Y. (2008): "The power of focal points is limited: even minute payoff asymmetry may yield large coordination failures", American Economic Review, vol. 98(4), pp. 1443– 58.

FISCHBACHER, U., (2007), "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." Experimental Economics, 10(2): 171–78.

GILBERT, M. (1989): On Social Facts. Routledge.

GOLD, N. (2012), "Team reasoning, framing and cooperation", in: S. OKASHA & K. B. BINMORE (Eds.), Evolution and rationality: Decision, cooperation and strategic behaviour. Cambridge: Cambridge University Press.

GOLD in Edited by Okasha, Samir and Binmore. Ken. Cambridge University Press, 2012
Evolution and Rationality: Decisions, Co-operation and Strategic Behaviour.

HARSANYI, J. and SELTEN, R. (1998): A general theory of equilibrium selection in games. Cambridge, MA: MIT Press.

HO, T., CAMERER C., WEIGELT K., (1998), "Iterated Dominance and Iterated Best Response in Experimental '**p**-Beauty Contests'." American Economic Review, 88(4): 947-969.

HODGSON, D. H. (1967): Consequences of Utilitarianism. Oxford: Clarendon Press.

HOLLIS, M. (1998): Trust within Reason. Cambridge: Cambridge University Press.

HURLEY, S. (1989): Natural Reasons. Oxford: Oxford University Press.

ISONI A. POULSEN A., SUGDEN R., TSUTSUI K., (2012), Focal points in tacit bargaining problems: Experimental evidence Volume 59, April 2013, 167‑188.

LEWIS, D.K. (1969): Convention: A Philosophical Study, Cambridge, MA: Harvard University Press.

MEHTA, J., STARMER, C. and SUGDEN, R. (1994): "The nature of salience: an experimental investigation", American Economic Review, vol. 84(3), pp. 658–73.

NAGEL, R. (1995): "Unraveling in guessing games: an experimental study", American Economic Review 85 (5), 1313–1326.

REGAN, D. (1980): Utilitarianism and Cooperation. Oxford: Clarendon Press.

SCHELLING, T. (1960): The strategy of conflict. Cambridge, MA: Harward University Press.

SMERILLI A. (2012) "We-thinking and vacillation between frames: filling a gap in Bacharach's theory", Theory and decision, 73(4), 539-560.

STAHL, D. O., WILSON, P., (1995): "Experimental evidence on players' models of other players", Journal of Economic Behavior and Organization 25 (3), 309–327.

STAHL, D. O., WILSON P., (1994): "Experimental Evidence on Players' Models of Other Players", Journal of Economic Behavior and Organization, 25 (3): 309-327.

SUGDEN, R. (2003): "The Logic of Team Reasoning," Philosophical explorations, 16, 165-181.

SUGDEN, R. (2000): "Team Preferences," Economics and Philosophy, 16, 175-204.

SUGDEN, R. (1993): "Thinking as a Team: Toward an Explanation of Nonselfish Behavior," Social Philosophy and Policy, 10, 69-89.

TUOMELA, R. (2007): The Philosophy of Social Practices: A Collective Acceptance View. Cambridge University Press.

TUOMELA, R. (1995): The Importance of Us: A Philosophical Study of Basic Social Notions. Stanford University Press.

TAN, J., and D. ZIZZO (2008): "Groups, Cooperation and Conflict in Games," The Journal of Socio-Economics, 37, 1-17.

FIGURES AND TABLES

Figure 1: Pie game's payoff matrix.

	R1	R2	R3
R1	x,y	0,0	0,0
R2	0,0	y,x	0,0
R3	0,0	0,0	v,w

Figure 2: The three treatments (Game G1)

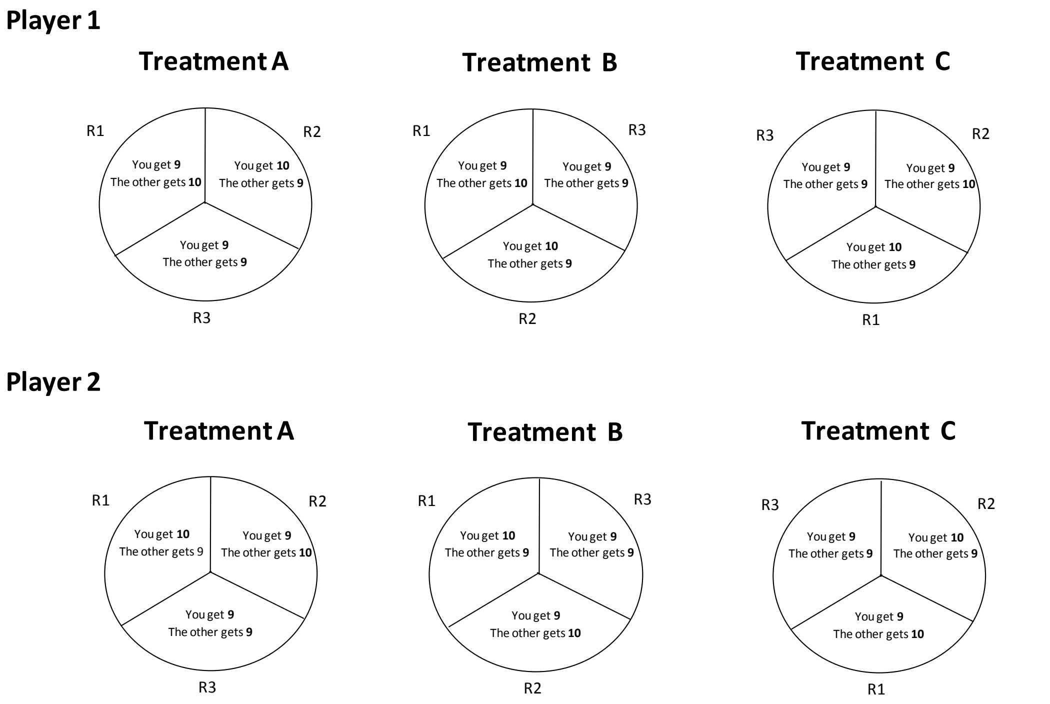
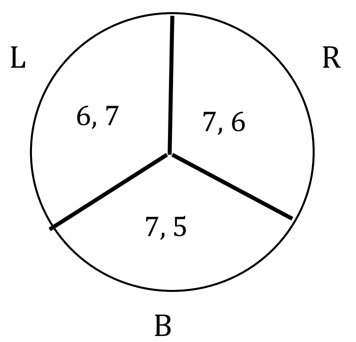


Figure 3. Crawford's et al. (2008) AM4 Game



TABLES

Table 1. Outcomes properties.

Game	Payoff R1	Payoff R2	Payoff R3	R1/R2	R3
G1	9,10	10,9	9,9	P	E
G2	9,10	10,9	11,11	-	E,P
G3	9,10	10,9	9,8	P	-
G4	9,10	10,9	11,10	-	P
G5	10,10	10,10	9,9	E,P	E
G6	10,10	10,10	11,11	E	E,P
G7	10,10	10,10	9,8	E,P	-
G8	10,10	10,10	11,10	E	P
G9	9,12	12,9	10,11	-	-
G10	10,10	10,10	11,9	E	-
G11	9,11	11,9	10,10	-	E

Table 2: Level-K theory's prediction in Game 1 when L0 plays at random.

GAME G1	L0	L1 (best reply to L0)	L2 (best reply to L1)	L3 (best reply to L2)
Player 1 choice	R1, R2, or R3	R2	R2	R1
Player 2 choice	R1, R2, or R3	R1	R1	R2

Table 3: Level-K theory's prediction in Game 1 when L0 follows payoff salience.

GAME G1	L0	L1 (best reply to L0)	L2 (best reply to L1)
Player 1 choice	R2	R1	R2
Player 2 choice	R1	R2	R1

Table 4. Predicted proportions of R3 choices.

		Predicted proportion of R3 choices			
Game	Team Reasoning	Level-k			
		L0	L1	L2	L3
G1	1	1/3	0	0	0
G2	1	1/3	1	1	1
G3	1	1/3	0	0	0
G4	1	1/3	3/4	1	1
G5	1	1/3	0	0	0
G6	1	1/3	1	1	1
G7	1	1/3	0	0	0
G8	1	1/3	2/3	1	1
G9	1	1/3	0	0	0
G10	1	1/3	1/2	1/2	1/2
G11	1	1/3	0	0	0

Table 5. Characteristics of games

Favoured choice	Criterion	Characteristics	GAMES										
R1, R2	b	R3 is Par. dominated by R1, R2	G1		G3		G5		G7				
	c	No PD									G9	G10	G11
	e	R3 unequal, R1, R2 equal							G7	G8		G10	
R3	a	R3 Par. dominates R1, R2		G2		G4		G6		G8			
	d	R3 equal, R1, R2 unequal	G1	G2									G11

Table 6. Distribution of choices in games G1-G11, coordination rates and predictions.

GAME	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
					Predicted R3 proportion						
Level-K	< 1/3	> 1/3	< 1/3	> 1/3	< 1/3	> 1/3	< 1/3	> 1/3	< 1/3	> 1/3	< 1/3
Team reasoning	> 1/3	> 1/3	> 1/3	> 1/3	> 1/3	> 1/3	> 1/3	> 1/3	> 1/3	> 1/3	> 1/3
					Frequencies R1						
P1	4	0	51	17	48	0	50	23	7	38	1
P2	16	0	48	15	46	1	49	28	24	45	8
TOT.R1	**20 (14.3%)**	**0 (0%)**	**99 (51%)**	**32 (16.5%)**	**94 (48.5%)**	**1 (0.5%)**	**99 (51.1%)**	**51 (26.3%)**	**31 (16%)**	**83 (42.8%)**	**9 (6.4%)**
					Frequencies R2						
P1	14	1	41	3	30	3	31	22	12	31	9
P2	2	0	47	4	36	3	30	21	10	21	1
TOT.R2	**16 (11.4%)**	**1 (0%)**	**88 (45.3%)**	**7 (3.7%)**	**66 (34%)**	**6 (3.1%)**	**61 (31.4%)**	**43 (22.2%)**	**22 (11.3%)**	**52 (26.8%)**	**10 (7.1%)**
					Frequencies R3						
P1	52	96	5	77	19	94	16	52	78	28	60
P2	52	97	2	78	15	93	18	48	63	31	61
TOT. R3	**104* (74.3%)**	**193* (99%)**	**7# (3.7%)**	**155* (79%)**	**34# (17.5%)**	**187* (96.4%)**	**34# (17.5%)**	**100* (51.5%)**	**141* (72.7%)**	**59 (30.4%)**	**121* (86.4%)**

* Proportion significantly (at 5%) higher than 1/3 (Two tails binomial test);
\# Proportion significantly (at 5%) lower than 1/3 (Two tails binomial test)

TOT	140	194	194	194	194	194	194	194	194	194	140
					Coordination rates						
	56%	99%	41.2%	71.1%	42.2%	92.8%	32%	42.2%	57.7%	32%	74.3%

Appendix 1. Games sequences.

Games sequences (20 subjects sessions)

	Round										
Subjects	1	2	3	4	5	6	7	8	9	10	11
1	1A(1)	2B(1)	3C(1)	4A(1)	5B(1)	6C(1)	7B(1)	8C(1)	9A(1)	10B(1)	11A(1)
2	11B(1)	1A(1)	2C(1)	3A(1)	4B(1)	10C(1)	6B(1)	7C(1)	8A(1)	9B(1)	5C(1)
3	4C(1)	5B(1)	11C(1)	2C(1)	3B(1)	9C(1)	10B(1)	6C(1)	7A(1)	8B(1)	1C(1)
4	3C(1)	4B(1)	5C(1)	1A(1)	2B(1)	8C(1)	9B(1)	10C(1)	6A(1)	7B(1)	11C(1)
5	2A(1)	3B(1)	4C(1)	5A(1)	11B(1)	7C(1)	8B(1)	9C(1)	10A(1)	6B(1)	1B(1)
6	1A(2)	3B(2)	5C(2)	2C(2)	4B(2)	6C(2)	8B(2)	10C(2)	7A(2)	9B(2)	11A
7	11B(2)	2B(2)	4C(2)	1A(2)	3B(2)	10C(2)	7B(2)	9C(2)	6A(2)	8B(2)	5C(2)
8	4C(2)	1A(2)	3C(2)	5A(2)	2B(2)	9C(2)	6B(2)	8C(2)	10A(2)	7B(2)	1C(2)
9	3C(2)	5B(2)	2C(2)	4A(2)	11B(2)	8C(2)	10B(2)	7C(2)	9A(2)	6B(2)	11C(2)
10	2A(2)	4B(2)	11C(2)	3A(2)	5B(2)	7C(2)	9B(2)	6C(2)	8A(2)	10B(2)	1B(2)
11	11A(1)	7B(1)	8C(1)	9A(1)	10B(1)	1A(1)	2B(1)	3C(1)	4A(1)	5B(1)	6A(1)
12	10A(1)	11B(1)	7C(1)	8A(1)	9B(1)	5A(1)	1B(1)	2C(1)	3A(1)	4B(1)	6B(1)
13	9A(1)	10B(1)	6C(1)	11C(1)	8B(1)	4A(1)	5B(1)	1C(1)	2A(1)	3B(1)	7A(1)
14	8A(1)	9B(1)	10C(1)	6A(1)	7B(1)	3A(1)	4B(1)	5C(1)	1B(1)	2B(1)	11A(1)
15	7A(1)	8B(1)	9C(1)	10A(1)	6B(1)	11B(1)	3B(1)	4C(1)	5A(1)	1C(1)	2A(1)
16	11A(2)	8B(2)	10C(2)	11C(2)	9B(2)	1A(2)	3B(2)	5C(2)	2A(2)	4B(2)	6A(2)
17	10A(2)	7B(2)	9C(2)	6A(2)	8B(2)	5A(2)	2B(2)	4C(2)	1B(2)	3B(2)	6B(2)
18	9A(2)	11B(2)	8C(2)	10A(2)	7B(2)	4A(2)	1B(2)	3C(2)	5A(2)	2B(2)	7A(2)
19	8A(2)	10B(2)	7C(2)	9A(2)	6B(2)	3A(2)	5B(2)	2C(2)	4A(2)	1C(2)	11A(2)
20	7A(2)	9B(2)	6C(2)	8A(2)	10B(2)	11B(2)	4B(2)	1C(2)	3A(2)	5B(2)	2A(2)

The first number indicates the game (from 1 to 11), the letter after the comma corresponds to the treatment (from A to B). The number in parentheses refers to the role (1=player 1; 2=player 2) under which the game is played.

Games sequences (18 subjects sessions)

	Round										
Subjects	1	2	3	4	5	6	7	8	9	10	11
1	1B(1)	2B(1)	3C(1)	4C(1)	5C(1)	6C(1)	7C(1)	8C(1)	9C(1)	10C(1)	11C(1)
2	11B(1)	1C(1)	2C(1)	3C(1)	4B(1)	10A(1)	6B(1)	7C(1)	8C(1)	9B(1)	5A(1)
3	4A(1)	11C(1)	3B(1)	1B(1)	2A(1)	9A(1)	10B(1)	8B(1)	6B(1)	7A(1)	5B(1)
4	3A(1)	4B(1)	5A(1)	2C(1)	1A(1)	8A(1)	9B(1)	10A(1)	7B(1)	6A(1)	11A(1)
5	1B(2)	4B(2)	3B(2)	3C(2)	2A(2)	6C(2)	9B(2)	10A(2)	8C(2)	7A(2)	11C(2)
6	11B(2)	2B(2)	5A(2)	1B(2)	4B(2)	10A(2)	7C(2)	8B(2)	6B(2)	9B(2)	5A(2)
7	4A(2)	1C(2)	3C(2)	2C(2)	5C(2)	9A(2)	6B(2)	8C(2)	7B(2)	10C(2)	5B(2)
8	3A(2)	11C(2)	2C(2)	4C(2)	1A(2)	8A(2)	10B(2)	7C(2)	9C(2)	6A(2)	11A(2)
9	11A(1)	7B(1)	8C(1)	9A(1)	10B(1)	1A(1)	2B(1)	3C(1)	4A(1)	5B(1)	6A(1)
10	10A(1)	11B(1)	7C(1)	8A(1)	9B(1)	5A(1)	1B(1)	2C(1)	3A(1)	4B(1)	6B(1)
11	9A(1)	10B(1)	6C(1)	11C(1)	8B(1)	4A(1)	5B(1)	1C(1)	2A(1)	3B(1)	7A(1)
12	8A(1)	9B(1)	10C(1)	6A(1)	7B(1)	3A(1)	4B(1)	5C(1)	1A(1)	2B(1)	11A(1)
13	7A(1)	8B(1)	9C(1)	10A(1)	6B(1)	11B(1)	3B(1)	4C(1)	5A(1)	1C(1)	2A(1)
14	11A(2)	8B(2)	10C(2)	11C(2)	9B(2)	1A(2)	3B(2)	5C(2)	2A(2)	4B(2)	6A(2)
15	10A(2)	7B(2)	9C(2)	6A(2)	8B(2)	5A(2)	2B(2)	4C(2)	1A(2)	3B(2)	6B(2)
16	9A(2)	11B(2)	8C(2)	10A(2)	7B(2)	4A(2)	1B(2)	3C(2)	5A(2)	2B(2)	7A(2)
17	8A(2)	10B(2)	7C(2)	9A(2)	6B(2)	3A(2)	5B(2)	2C(2)	4A(2)	1C(2)	11A(2)
18	7A(2)	9B(2)	6C(2)	8A(2)	10B(2)	11B(2)	4B(2)	1C(2)	3A(2)	5B(2)	2A(2)

The first number indicates the game (from 1 to 11), the letter after the comma corresponds to the treatment (from A to B). The number in parentheses refers to the role (1=player 1; 2=player 2) under which the game is played.

Games sequences (16 subjects sessions)

	Round										
Subjects	1	2	3	4	5	6	7	8	9	10	11
1	11A(1)	2B(1)	3C(1)	4C(1)	5C(1)	6C(1)	7B(1)	8C(1)	9C(1)	10C(1)	1C(1)
2	5A(1)	1B(1)	2C(1)	3C(1)	4B(1)	10A(1)	6B(1)	7C(1)	8C(1)	9B(1)	11B(1)
3	4A(1)	11C(1)	1A(1)	5B(1)	2A(1)	9A(1)	8B(1)	6A(1)	10B(1)	7B(1)	3B(1)
4	11A(2)	11C(2)	2C(2)	4C(2)	5C(2)	6C(2)	8B(2)	7C(2)	9C(2)	10C(2)	1C(2)
5	5A(2)	2B(2)	1A(2)	3C(2)	4B(2)	10A(2)	7B(2)	6A(2)	8C(2)	9B(2)	11B(2)
6	4A(2)	1B(2)	3C(2)	5B(2)	2A(2)	9A(2)	6B(2)	8C(2)	10B(2)	7B(2)	3B(2)
7	11A(1)	7B(1)	8C(1)	9A(1)	10B(1)	1A(1)	2B(1)	3C(1)	4A(1)	5B(1)	6A(1)
8	10A(1)	11B(1)	7C(1)	8A(1)	9B(1)	5A(1)	1B(1)	2C(1)	3A(1)	4B(1)	6B(1)
9	9A(1)	10B(1)	6C(1)	11C(1)	8B(1)	4A(1)	5B(1)	1C(1)	2A(1)	3B(1)	7A(1)
10	8A(1)	9B(1)	10C(1)	6A(1)	7B(1)	3A(1)	4B(1)	5C(1)	1A(1)	2B(1)	11A(1)
11	7A(1)	8B(1)	9C(1)	10A(1)	6B(1)	11B(1)	3B(1)	4C(1)	5A(1)	1B(1)	2A(1)
12	11A(2)	8B(2)	10C(2)	11C(2)	9B(2)	1A(2)	3B(2)	5C(2)	2A(2)	4B(2)	6A(2)
13	10A(2)	7B(2)	9C(2)	6A(2)	8B(2)	5A(2)	2B(2)	4C(2)	1A(2)	3B(2)	6B(2)
14	9A(2)	11B(2)	8C(2)	10A(2)	7B(2)	4A(2)	1B(2)	3C(2)	5A(2)	2B(2)	7A(2)
15	8A(2)	10B(2)	7C(2)	9A(2)	6B(2)	3A(2)	5B(2)	2C(2)	4A(2)	1B(2)	11A(2)
16	7A(2)	9B(2)	6C(2)	8A(2)	10B(2)	11B(2)	4B(2)	1C(2)	3A(2)	5B(2)	2A(2)

The first number indicates the game (from 1 to 11), the letter after the comma corresponds to the treatment (from A to B). The number in parentheses refers to the role (1=player 1; 2=player 2) under which the game is played.

Appendix 2. Distribution of choices across treatments

GAME G1		Treat.A	Treat.B	Treat.C
P1	R1/R2	3	7	8
	R3	25	14	13
	TOT	**28**	**21**	**21**
	Binomial two sided test (n=R1+R2+R3 choices; k=R3 choices; p=0.33)	p=0.000	p=0.001	p=0.008
	H0 : same distr. across treatments ; Pearson chi2(2) = 5.62 p>0.05			
P2	R1/R2	5	5	8
	R3	23	16	13
	TOT	**28**	**21**	**21**
	Binomial two sided test (n=R1+R2+R3 choices; k=R3 choices; p=0.33)	p=0.000	p=0.000	p=0.008
	H0 : same distr. across treatments ; Pearson chi2(2) = 2.63 p>0.05			
	Coord. Rate (%)	75	52.3	38

GAME G2		Treat.A	Treat.B	Treat.C
P1	R1/R2	1	0	0
	R3	31	39	26
	TOT	**32**	**39**	**26**
	Binomial t	p=0.000	p=0.000	p=0.000
	Pearson chi2(2) = 2.0524 p>0.05			
P2	R1/R2	0	0	0
	R3	32	39	26
	TOT	**32**	**39**	**26**
	Binomial t	p=0.000	p=0.000	p=0.000
	Pearson chi2(2) = -			
	Coord. Rat	96.9	100	100

GAME G3		Treat.A	Treat.B	Treat.C
P1	R1/R2	29	37	26
	R3	2	2	1
	TOT	**31**	**39**	**27**
	Binomial two sided test (n=R1+R2+R3 choices; k=R3 choices; p=0.33)	p=0.000	p=0.000	p=0.000
	Pearson chi2(2) = 0.22 p>0.05			
P2	R1/R2	30	38	27
	R3	1	1	0
	TOT	**31**	**39**	**27**
	Binomial two sided test (n=R1+R2+R3 choices; k=R3 choices; p=0.33)	p=0.000	p=0.000	p=0.000
	Pearson chi2(2) = 0.82 p>0.05			
	Coord. Rate (%)	45	33	48

GAME G4		Treat.A	Treat.B	Treat.C
P1	R1/R2	6	10	4
	R2			
	R3	27	29	21
	TOT	**33**	**39**	**25**
	Binomial t	p=0.000	p=0.000	p=0.000
	Pearson chi2(2) = 1.04 p>0.05			
P2	R1/R2	8	5	6
	R3	25	34	19
	TOT	**33**	**39**	**25**
	Binomial t	p=0.000	p=0.000	p=0.000
	Pearson chi2(2) = 1.89 p>0.05			
	Coord. Rat	75.7	69.2	68

GAME G5	Treat.A	Treat.B	Treat.C
R1	28	30	20
R2			
P1 R3	6	8	5
TOT	**34**	**38**	**25**
Binomial two sided test (n=R1+R2+R3 choices; k=R3 choices; p=0.33)	p=0.23	p=0.12	p=0.20
H0 : same distr. across treatments; Pearson chi2(2) = 0.13 p>0.05			
R1/R2	31	31	20
P2 R3	3	7	5
TOT	**34**	**38**	**25**
Binomial two sided test (n=R1+R2+R3 choices; k=R3 choices; p=0.33)	p=0.001	p=0.06	p=0.20
H0 : same distr. across treatments; Pearson chi2(2) = 1.79 p>0.05			
Coord. Rate (%)	44	42	40

GAME G6	Treat.A	Treat.B	Treat.C
R1/R2	0	0	3
P1 R3	33	39	22
TOT	**33**	**39**	**25**
Binomial t	p=0.000	p=0.000	p=0.000
Pearson chi2(2) = 8.9157 p<0.05			
R1/R2	1	1	2
P2 R3	32	38	23
TOT	**33**	**39**	**25**
Binomial t	p=0.000	p=0.000	p=0.000
Pearson chi2(2) = 1.28 p>0.05			
Coord. Ra	97	97.4	80

GAME G7	Treat.A	Treat.B	Treat.C
R1/R2	24	35	22
P1 R3	7	5	4
TOT	**31**	**40**	**26**
Binomial two sided test (n=R1+R2+R3 choices; k=R3 choices; p=0.33)	p=0.25	p=0.004	p=0.06
Pearson chi2(2) = 1.32 p>0.05			
R1/R2	25	33	21
P2 R3	6	7	5
TOT	**31**	**40**	**26**
Binomial two sided test (n=R1+R2+R3 choices; k=R3 choices; p=0.33)	p=0.12	p=0.04	p=0.15
Pearson chi2(2) = 0.05 p>0.05			
Coord. Rate (%)	32.3	40	19.2

GAME G8	Treat.A	Treat.B	Treat.C
R1/R2	17	16	12
P1 R3	14	23	15
TOT	**31**	**39**	**27**
Binomial t	p=0.18	p=0.001	p=0.02
Pearson chi2(2) = 1.38 p>0.05			
R1/R2	16	21	12
P2 R3	15	18	15
TOT	**31**	**39**	**27**
Binomial t	p=0.08	p=0.09	p=0.02
Pearson chi2(2) = 0.58 p>0.05			
Coord. Ra	37	43.5	44.4

GAME G9		Treat.A	Treat.B	Treat.C
	R1/R2	8	5	6
P1				
	R3	25	34	19
	TOT	**33**	**39**	**25**
	Binomial two sided test (n=R1+R2+R3 choices; k=R3 choices; p=0.33)	p<0.01	p<0.01	p<0.01
	H0 : same distr. across treatments; Pearson chi2(2) = 1.89 p>0.05			
	R1/R2	10	10	14
P2				
	R3	23	29	11
	TOT	**33**	**39**	**25**
	Binomial two sided test (n=R1+R2+R3 choices; k=R3 choices; p=0.33)	p=0.000	p=0.000	p=0.28
	H0 : same distr. across treatments; Pearson chi2(4) = 6.66 p<0.05			
	Coord. Rate (%)	60.6	69	36

GAME G10		Treat.A	Treat.B	Treat.C
	R1/R2	21	28	20
P1				
	R3	13	10	5
	TOT	**34**	**38**	**25**
	Binomial t	p=0.58	p=0.49	p=0.20
	Pearson chi2(2) = 1.18 p>0.05			
	R1/R2	21	27	18
P2				
	R3	13	11	7
	TOT	**34**	**38**	**25**
	Binomial t	p=0.58	p=0.73	p=0.67
	Pearson chi2(2) = 0.95 p>0.05			
	Coord. Rat	26.4	34.2	36

GAME G11		Treat.A	Treat.B	Treat.C
	R1/R2	2	6	2
P1				
	R3	19	22	19
	TOT	**21**	**28**	**21**
	Binomial two sided test (n=R1+R2+R3 choices; k=R3 choices; p=0.33)	p=0.000	p=0.000	p=0.000
	Pearson chi2(2) = 1.94 p>0.05			
	R1/R2	3	2	4
P2				
	R3	18	26	17
	TOT	**21**	**28**	**21**
	Binomial two sided test (n=R1+R2+R3 choices; k=R3 choices; p=0.33)	p=0.000	p=0.000	p=0.000
	Pearson chi2(2) = 1.57 p>0.05			
	Coord. Rate (%)	76.2	71.4	76.2