NEW PERSPECTIVES ON COOPERATION AND TEAM REASONING: THEORY AND EXPERIMENTS

Alessandra Smerilli

Submitted for the degree of Doctor of Philosophy

University of East Anglia
School of Economics
July 2014-07-23

ABSTRACT

Players' use of cooperative strategies in Prisoner's Dilemma (PD) games and their achievement of coordination in some kinds of coordination games are among the most studied issues in both theoretical and experimental game theory. The present thesis is a collection of three article on this topic.

Chapter 2 of the thesis focuses on cooperation, by developing an evolutionary model of a repeated Prisoner's Dilemma game, using replicator dynamics. The evolution of cooperation is analysed in terms of the interaction of different strategies, which represent the heterogeneity of forms of cooperation in civil life. One of the results of the paper is the conclusion that cooperation is favoured by heterogeneity: the presence of different kinds of strategies enhances cooperation.

A theory that can explain both cooperation and coordination is team reasoning. Chapter 3 represents a development of Bacharach's theory of team reasoning. Starting from a detailed review of Bacharach's writings, and in order to clarify some issues linked to reasoning and frames, I propose a 'vacillation' model in which agents are allowed to have both I and we- concepts in their frames, and can easily switch from one to another.

The theoretical model presented in Chapter 3 is followed by an experiment, reported in Chapter 4. The experiment aims at identifying which features of the structure of payoffs in coordination games favour the use of team reasoning, using Level-k theories as the benchmark for the modelling of individual reasoning. We find mixed evidence about level-k and team reasoning theories. In particular team reasoning theory fails to predict choices when it picks out a solution which is Pareto dominated and not compensated by greater equality. This could represent a step forward in investigating the roles of team reasoning and level-k reasoning in explaining coordinating behaviour.

ACKNOWLEDGMENTS

I would like to take this opportunity to thank all the persons who have contributed in many respects to this research.

First of all I'm deeply grateful to my primary supervisor, Robert Sugden, for his expert support and encouragement during the whole period of study.

I'm also grateful to the second supervisor, Shaun Hargreaves Heap, with whom I discussed chapter 3 of the thesis.

I would like to thank the two examiners, Natalie Gold and Ben McQuillin for reading careflully my thesis and for the useful comments on it.

Many thanks also to all the members of the School of Economics and to my colleagues.

I wish to thank my Institution, The PFSE-Auxilium in Rome, and my religious community for letting me study abroad.

During these years I have been very lucky to meet people who have greatly enriched the quality of my research. I would like to thank Federica Alberti, Luigino Bruni, Luca Crivelli, Marco Faillo, Benedetto Gui, Vittorio Pelligra, Pier Luigi Porta, Kei Tsutsui and Stefano Zamagni.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

The three articles collected in this thesis include some collaborative work. Chapter 2 is a joint work with Luigino Bruni. The mathematical evolutionary model – including the proofs in the Appendix -, is entirely my work and I also contributed to the writing up. The paper has been published in *Homo Oeconomicus* (vol. 29 (2) 2012).

The content of Chapter 3 is entirely my work, carried out under Robert Sugden's supervision. I have benefited from comments by Shaun Heargreaves Heap in his role as my second supervisor, and by two anonymous referees. This chapter has been published in *Theory and Decision* (vol. 73 (4)2012).

The experiment reported in chapter 4 is the result of collaboration with Robert Sugden and Marco Faillo. I contributed to this paper with the experimental design, which has been done under the supervision of Robert Sugden, some of the data analysis and the writing up the results in the current form. Marco Faillo is responsible for programming the experiment with z-three software and for some of the data analysis. The experiment was run by Marco Faillo and me.

Chapter 1.
Introduction

"It is hard to resist the understanding that social identity

is a significant influence on human behaviour.

The idea that a sense of community and fellowship is important for us all

is also difficult to ignore, and it relates closely to our conception of social identity".

(Sen 1999, p. 5)

1. Cooperation and coordination in game theory: main issues

Players' use of cooperative strategies in Prisoner's Dilemma (PD) games and their achievement of coordination in some kinds of coordination games are among the most studied issues in both theoretical and experimental game theory.

In classical game theory there are problems in explaining both cooperation and coordination. The problem of coordination in games arises from the existence of multiple equilibria. In one-shot coordination games, classical game theory cannot establish which equilibrium will be selected. According to Colin Camerer: "Predicting which of many equilibria will be selected is perhaps the most difficult problem in game theory. This selection problem is essentially unsolved by analytical theory" (2003, p. 336).

Some refinements of equilibrium, such as payoff dominance and risk dominance (Harsanyi and Selten 1988), have been proposed as means of solving this problem. However, such refinements need a broader theory to explain why rational agents can be guided by the principles on which those school was searching for": a payoff-dominated Nash equilibrium "is not only an equilibrium; it is a perfectly good equilibrium". Another device which helps to choice among equilibria in coordination problem is 'focality', linked to the so called 'Schelling salience': sometimes equilibria are focal or psychologically prominent (Schelling,1960; Metha et al., 1994; Bardsley et al., 2010; Crawford et al., 2008; Isoni et al., 2012). But game theory needs to explain *why* rational players are influenced by focality.

One-shot PD games pose a different problem. In a PD game there is a unique Nash equilibrium – defection by both players – which is Pareto dominated by the combination of payoffs resulting from cooperation. Classical game theory implies that rational players would always defect. However, experimental evidence shows that subjects cooperate much more than standard game theory predicts[1]. Thousands of pages have been written on how cooperative choices in a PD game can be explained. Some theorists have explained cooperation in terms of adherence to moral rules grounded on Kantian morality (Laffont, 1975; Collard, 1978, 1983), reciprocity (Sugden, 1984), instrumental rationality (Gauthier, 1986), or expressive rationality (Benn, 1978; Hargreaves Heap, 1997), or to social norms (Elster, 1989; Bicchieri, 1997, 2006). Other theorists have explained cooperation as resulting from social preferences reflecting altruism (Sober and Wilson 1988, Bergstrom and Stark, 1993,Taylor 1976, 1987), inequity aversion (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999), reciprocity and strong reciprocity (Axelrod 1984; Rabin 1993; Levine, 1998; Falk and Fischbacher, 2005; Dufwenberg and Kirchsteiger, 2004; Segal and Sobel, 1999; Fehr and Fiscbacher, 2005).

A theory that can explain both cooperation and coordination is team reasoning. Chapters 3 and 4 of this thesis are concerned with team reasoning. I shall introduce team reasoning theories later on in the introduction, by presenting a brief literature review.

Another branch of literature that deals with cooperation and coordination in game theory is evolutionary game theory. Chapter 2 of this thesis reports a

[1] See Hargreaves Heap and Varoufakis (2004) for a review.

theoretical model which makes use of evolutionary game theory in order to explain cooperation in a PD game. I will introduce it in the next section.

2. Evolution of cooperation

Evolutionary game theory, in its best known form, is based on the evolution of proportions of relevant strategies present in a population, according to replicator dynamics, which embodies the idea of Darwinian selection, or of the imitation of the most successful strategies over time.

Although evolutionary ideas have a long history in economics (Darwin 1859 referred to Malthus's theory, Marshall [1890, xiv] claims that economic biology is the 'Mecca' of economists), the use of them in game theory is relatively recent (see Boyd and Richerson, 1985, Sugden 2004 [first edition 1986], Weibull, 1995). Evolutionary game theory offers explanations of coordination (see Balkenborg, 1993 and Robles 2001) and cooperation (see Axelrod 1984, Hoffman 1999, Sugden 2004).

Chapter 2 of the thesis focuses on cooperation, by developing an evolutionary model of a repeated Prisoner's Dilemma game, using replicator dynamics. The evolution of cooperation is analysed in terms of the interaction of different strategies, which represent the heterogeneity of forms of cooperation in civil life. The model uses four strategies, adapted from Sugden's model of reciprocity (2004): N (always defect), B (a trigger strategy which begins by cooperating), C (a trigger strategy which begins by defection) and G (always cooperate).

One of the results of the paper is the conclusion that cooperation is favoured by heterogeneity: the presence of different kinds of strategies enhances cooperation. More to the point, the model shows that unconditional actions of cooperation (the G strategy) can be essential if cooperation is to establish itself in populations initially characterised by generalised non-cooperation; unconditional cooperation acts as a starter. But, at the same time, unconditional actions of cooperation should not be too frequent, otherwise they become counter-productive. Although the latter result is shared with Hoffmann (1999), who runs simulations and finds a turnover between cooperative stages and non-cooperative ones, and with Sugden (2004), who explains this turnover by means of 'sleeper' strategies[2], nevertheless the role of G strategies is underestimated in the previous models.

Usually G strategies are considered useless or even detrimental for the emergence of cooperation: in most of the relevant literature they are called 'sucker' strategies. (For example, Heller and Sieberg (2010) argue that cooperation can survive without the existence of unconditional co-operators). In our model they are essential because they are the only kind of strategies which can activate C. In Sugden's model, 'cautious reciprocator' strategies (analogous with C) can be activated to cooperate by 'brave reciprocator' strategies (analogous with B), but only because 'caution' takes a sophisticated form, tailored to the specific punishment behaviour of the brave strategy. But the point of view of Chapter 2 is different: our analysis starts from the assumption that G behaviour does exist in society (even Binmore, 2006 admits its existence). We try to demonstrate that unconditional cooperation can not only survive, but can also perform a role in fostering cooperation.

3. Team reasoning: a theory that can explain cooperation and coordination

Theories of team reasoning have been proposed as explanations of cooperation and coordination in games. 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.

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:

[2] Sleeper strategies are strategies whose behaviour is indistinguishable from that of the incumbent strategy, but which favour certain kinds of invaders if and when they arrive. In this case the sleeper strategy is 'cautious reciprocity'.

"What should we do?", and they act accordingly. The use of team reasoning can explain both cooperation in a PD game and successful coordination in coordination games.

All the existing theories of we-thinking share the idea that each member of a set of individuals who are engaged in a strategic interaction can 'identify' with that set as a 'group'. Each member of the group then tries to reach the best outcome for the whole group, doing his/her part in the best combination of actions. This kind of reasoning is called team or we-reasoning and it is the effect of group identification.

3.1 Differences in theoretical approaches

Theories of we-reasoning are basically divided into two kinds of approaches: some scholars consider group formation as rational, others do not. For Bacharach, for example, team reasoning is a result of a psychological mechanism – group identification -; for Gilbert (1989) and Tuomela (1995) group formation is a result of a mutual commitment; Regan, in his *Utilitarianism and Co-operation* (1980) proposes a normative theory for moral and rational agents in which the rule to follow is: "What each agent ought to do is to cooperate, with whoever else is cooperating, in the production of the best consequences possible given the behaviour of non cooperators" (p.124).

Susan Hurley offers a theory of 'rationality' of we-thinking. Defining the unit of agency as "the unit the causal consequences of the activity of which are in question" (1989, p.140), Hurley identifies as units the subsystem ('each') or the system ('we') and claims that those units have not to be taken as fixed. In fact, facing a decision problem an agent firstly must ask herself: which is the objective in this situation? Subsequently she can choose the unit of agency that is the most appropriate for the objective: "An adequate theory should help us to understand what the appropriate unit of agency is in various circumstances" (p.146). The consequence of this analysis is that it is rational to allow different units of agency, hence I-thinking or we-thinking can be rational, depending on the circumstances.

A question, however, remains open: "If units of agency are not exogenously fixed, how are units formed and selected? Is centralized information or control required, or can units emerge as needed from local interactions? At what points are unit formation and selection rationally assessable?" (2003, p.165). The last question is very important, because the fact that unit formation is rational has still to be demonstrated. Hollis and Sugden argue that, in answering the question whether the formation of a unit is a requirement of rationality: "If we are to stay at all close to account of rationality that derives from Hobbes, Hume, Bentham, Pareto and Savage, we must answer 'No' (Hollis and Sugden 1993, p. 13)". According to their account of rational choice theory, a choice is rational in relation to the desires or preferences of the agent who is making the choice: "a choice can be rational only for a particular agent" (ib.). It follows that a theory of rationality cannot give an account of the formation of the unit of agency.

Differently from Hurley, who claims that there must be agent-neutral goals to be pursued, Elizabeth Anderson (2001) takes a position more similar to that of Hollis and Sugden. She states that the determination of personal identity, which can be plural or individualistic, precedes the choice of the kind of reasoning to be adopted. She makes use of team reasoning in order to give an account of the 'rationality of committed action': "regarding themselves as members of a single collective agency, the parties are committed to acting only on reasons that are universalizable to their membership"(p.29). She then states the 'Priority of Identity to Rational Principle': "what principle of choice it is rational to act on depends on a prior determination of personal identity, of who one is" (p.30). Following the previous principle, Anderson shows that either acting on maximization of expected utility or on team reasoning is a rational act, depending on regarding oneself as an isolated individual or a member of a team. In Anderson's account, then, the determination of personal identity comes before the decision of what principle of choice is in play.

Michael Bacharach offers the most developed theory of group identification. Like Hollis and Sugden, Bacharach treats agency as prior to rationality. His theory 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” (2006. p. 69). In Bacharach’s framework, then, a person may start to we-reason only if she has ‘we’ concepts in her frame: in other words, a person firstly recognizes the we-perspective, and then endorses it.

Robert Sugden has developed a different framework for looking at the problem. In his framework, the central concept is ‘common reason to believe’: people who group identify are not committed to reason as a team unless there is a common reason to believe that other agents are doing the same. The psychological side of group identification in Sugden’s theory might be found in his analysis of Smith’s ‘correspondence of sentiments’ (Sugden, R. 2005): ‘fellow-feeling’ could be seen as the source of group identification. To summarize, in Bacharach’s framework if people group identify they automatically start to reason like a team, whereas in Sugden’s theory people may group identify, but team reasoning does not follow automatically.

3.2 The I-we equilibrium

My approach to we-thinking, which is explained in chapter 3, is based on a development of Bacharach’s theory.

Starting from a detailed review of Bacharach’s writings, I aim to clarify some issues linked to reasoning and frames, and then to build a framework in which agents are allowed to have both I and we- concepts in their frames, and can easily switch from one to another.

By studying Bacharach’s published and unpublished papers, I identify a limitation in his theory: Bacharach cannot allow people to use more than one frame at a time. In a certain sense, as it has been noticed by Gold and Sugden (Bacharach, Gold and Sugden 2006), in the ‘we’ frame people become committed to we-reason: “In the theory of team reasoning, an individual who reasons in the ‘we’ frame is aware of the ‘I’ frame too (as one that other players might use) but acknowledges only ‘we’ reasons” (p.199).

The approach I follow shares with Hurley’s theory the idea that the unit of agency cannot be considered fixed, but not the idea that there must be agent-neutral goals to be pursued: in my account the agent does not rationally choose the unit of agency, but, with the help of reasoning, he/she can *vacillate* between units. The notion of vacillation between frames is based on Bacharach’s idea that frames are non-integrable and can not be present at the same time in a person. The image Bacharach uses to explain this concept is the famous Rubin’s vase. In my account an agent starts with one unit of agency (I or we) – which one it is depends on psychological considerations or framing matters, or salience… - and evaluates what is the best to do, given that agency.

In particular, if the agent starts by I-reasoning, he/she asks whether ‘we’ would be better off by deviating from I-reasoning. If, instead the agent starts by we-reasoning, he/she asks if ‘I’ would be better off by deviating from we-reasoning. With the help of this process, and by defining an equilibrium as a situation in which there is no incentive to deviate, I can identify I-equilibria (from which ‘I’ have no incentive to deviate), We-equilibra (from which ‘we’ have no incentive to deviate), and I-We equilibria, which are consistent with both I and we reasoning. In my account, then, as in Bacharach’s account there exist only I-reasoning and we-reasoning, without a higher-order reasoning. But, differently from Bacharach, I allow reasoning to have some influence on which unit of agency is being used.

4. An experiment

In explaining group identification, Bacharach presents his *interdependence hypothesis*, which is based, among other things, on the payoff structure of the game being played. This means that, according to Bacharach, some feature of the games can enhance team reasoning.

In order to clarify this issue, the theoretical model presented in Chapter 3 is followed by an experiment, reported in Chapter 4.

The experiment aims at identifying which features of the structure of payoffs in coordination games favour the use of team reasoning, using Level-k theories[3] as the benchmark for the modelling of individual reasoning. Level-k theories explain coordination in games avoiding the equilibrium selection problem: they anchors players' beliefs on the behaviour of strategically naïve individuals, who are supposed to play randomly or to choose payoff salient strategies.

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 others (e.g. Crawford et al., 2008; Bardsley et al., 2010; Isoni et al., 2012). Sometimes it is impossible to discriminate between them.

A way to explain these mixed results is to build a general theory of group identification, which could explain why group identification occurs in some games, in which the theory of team reasoning works, but not in other games. Bacharach (2006), Crawford et al. (2008) and Bardsley et al. (2010) offer conjectures about this, but they do not propose clear hypotheses.

In order to go deeper into the matter, I present an experiment using 'pie games', similar to those used by Crawford et al. (2008). Pie games are two-player coordination games in which all payoffs are zero except on the main diagonal; payoffs on the diagonal can differ both between players and between equilibria. In my experiment, each game has three strategies for each player and therefore three pure-strategy equilibria. Two of these equilibria have symmetrical payoffs (i.e. if one equilibrium has the payoffs (x, y), the other has (y, x)). The third equilibrium, which I call 'unique', has distinct payoffs.

In the games two characteristics, Pareto dominance and equality are present in various combinations, and the equilibrium with distinct payoffs is the unique 'we-solution' for each game. In particular, there are some games in which the we-reasoning solution Pareto-dominates the other two equilibria ex post (i.e. given that both players have coordinated on it), and there are other games in which the we-solution is Pareto-dominated by the other equilibria ex post. (In

[3] These theories will be explained in chapter 4.

all our games the we-solution Pareto-dominates the other choices ex ante – i.e. on the assumption that the players cannot coordinate on any one of the two symmetrical equilibria). Secondly, sometimes the we-solution is equal, or more equal than the other two; sometimes it is unequal, or more unequal than the others.

The subjects in the experiment play 11 games, representing all possible combinations among equality and Pareto-dominance. The aim is to examine how these features work in predicting choices and in prompting team reasoning. Our conjectures, related to the characteristics of the games, are that, having the we-solution as an equilibrium which is Pareto-dominated ex post by the others might discourage team reasoning, and the same will happen when the we-solution is unequal or more unequal compared to the other equilibria. Conversely, if the we-solution Pareto-dominates the others ex post, or if it is equal or more equal than the other equilibria, this condition will enhance team reasoning.

In the experiment there are 6 games in which team reasoning and level-k predictions disagree.

Our results confirm our conjectures, showing that Pareto dominance and equality are good predictors for coordination choices. We find mixed evidence about level-k and team reasoning theories. In particular 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 the choice of one non-unique equilibrium, and the unique equilibrium is more equal than the alternatives. This could represent a step forward in investigating the roles of team reasoning and level-k reasoning in explaining coordinating behaviour.

5. General conclusions from my work

At the end of this long path, I think is important to summarize what I have learned from my work.

First of all I have learned that, in order to foster cooperation, it is important to take into account the role of unconditionally cooperative strategies, and not to underestimate them. Civil life teaches us that sometimes it is impossible for cooperation to be sustained unless some people are willing to keep cooperating even when no one else does. Chapter 2 is an attempt to show how this observation can be represented in game theory.

I believe I have made a contribution to the foundations of team reasoning theory by introducing the possibility that a person can switch from I-reasoning to we-reasoning and vice versa, guided by considerations of rationality (reasoning about the incentives to deviate from one unit of agency to another) and not only by psychology. During the reviewing process for publishing Chapter 2 I learned that team reasoning is something so obvious and at the same time so strange for mainstream economics that it is easily misunderstood. More effort is needed in order to build a more general theory of team reasoning. This could involve enlarging the range of games to which the concept of I-we equilibrium can be applied, included dynamic games.

The experiment shows that team reasoning does exist, and that it is not so 'exotic' as Gintis (2003) claims. It shows also that sometimes team reasoning is used, sometimes not. I have made some progress in identifying conditions, linked to the payoff structure of games, which favour the use of team reasoning in coordination games. A possible line of development could be to use games in which it is possible to distinguish between strict and weak Pareto dominance, in order to study in a better way the effect of Pareto dominance on team reasoning.

Another useful development might be to use eye tracker tools in experiments, in order to identify the patterns of reasoning used by the subjects. In particular, Bacharach (2006) claims that team reasoners compare the different couples of payoffs (my payoff and my partner's one). An I-reasoner, instead, tries to find his best responses to his co-player's strategies, and to find *her* best responses to *his* strategies. He has no need to compare his payoffs with those of the co-player. Thus a Row player makes comparisons between his payoffs along each given row, and between the co-players' payoffs down each given column. Since the two kinds of reasoning involve different payoff comparisons, they might induce different patterns of eye movement.

Chapter 2:
Cooperation and diversity

1. Introduction

Civil life is essentially cooperation. Neoclassical economics offers mostly a parsimonious view of cooperation merely based on individual self-interests and instrumental rationality. In such a vision of cooperation, an agent will never cooperate in a one-shot Prisoner dilemma game. If instead the game is repeated, the traditional theory justifies the cooperation by evoking self-interest and/or enforcement (Binmore 2005, 2006). In reaction to this parsimonious view of cooperation, recent years have seen development of a body of literature (mainly experimental), the so-called 'social preferences' theories, which instead seeks to explain why even in a one-shot non-cooperative game (i.e. the 'ultimatum' or 'trust' game) it may be rational to play 'cooperatively'. The explanation, of which there are several variants, is a re-definition of the utility function of the agents, by introducing non material payoffs associated to norms such as inequality aversion or reciprocity. In this way it is possible to explain the emergence of cooperative behaviour in contexts where the standard theory would exclude it. This is the explanation of cooperation advanced by behavioural economists (see Gintis (2004) and Bowles and Gintis (2004)), who base their analyses of cooperation on the theory of strong reciprocity (Fehr and Gächter (2000)). By 'strong reciprocity' they mean a social norm which rewards those who behave in a kind way and punishes those who behave in an unkind way. This theory explains the emergence of cooperation on the basis of a form of altruism, which does not require the repetition of the game. In this paper we adopt a different approach for explaining the emergence of cooperation in an evolutionary scenario. We propose a pluralistic and multidimensional view of cooperation and consequently examine aspects hitherto not sufficiently explored by economic and social theory. In particular, the intuition inspiring this paper is the multidimensional nature of cooperation: civil society flourishes if and when different forms of reciprocity are seen as complementary rather than rival or substitute one another. In this sense, we shall show that diversity fosters cooperation, a result well known in biology. Specifically, we claim to show on the one hand that, in certain settings, less 'altruistic', conditional forms of cooperation may combine with unconditional ones generating a co- operative environment. On the other hand, we'll demonstrate that too many unconditional actions will end to promote the non-cooperation. The combination of these two results embodies the main contribution of the paper. We accordingly construct models of evolutionary game theory, which will enable us to analyse diverse patterns of cooperation, not all of them based on self-interest, but all of them important for understanding the dynamics of civil life. We shall base our analysis on dynamic Prisoner's Dilemma (PD) game, because it lends itself well to the modelling of 'difficult' cooperation, the kind that occurs in situations where there is no enforcement and where there is always an incentive for non-cooperation. We believe that these situations are frequent and relevant – although in civil society individuals play many games, not only the PD – and that they are important in the real dynamics of cooperation in civil life. In section 2 we introduce our model. In section 3 we analyse the evolution of cooperation in a context of repeated games. In section 4 we concentrate on analysis of situations in which four strategies interact, also making some simulations. The paper concludes with a brief discussion of the results of our analysis.

2. The repeated dynamic game

The pay-off matrix of the game is the following.[4]

Table 2.1: Prisoner's dilemma

	C	**D**
C	β- γ	- γ
D	β	0

[4] The table represents a particular case which simplifies the analysis without compromising the results. As well known, for a game to be a Prisoner's Dilemma, the payoff order must be $\beta > \gamma > 0$.

It can be easily shown that both players will choose not to cooperate or defect (D) in a one-shot game, and that the outcome (0,0) will be a Nash equilibrium.[5] In this kind of non-iterated game cooperation cannot arise unless errors are committed or the players behave irrationally.

Our model makes use of repeated games. We assume, that is to say, that associated with every random encounter is a repeated interaction with the same person. This interaction may be of greater or lesser duration according to a parameter, π, which denotes the discount factor[6]. We are hence in a context of indefinitely repeated game. After a series of interactions with the same person, another random encounter occurs, and the (repeated) game resumes with another (randomly matched) partner.

The structure of our model is as follows. Time is continuous. We suppose that there is a continuum of agents belonging to a particular population, and that they must choose one of the J pure strategies $\{1,\ldots,J\}$ whenever they interact with other subjects in the same population. The subjects are distributed among I sub-populations $\{1,\ldots,I\}$, which are assigned exogenously in the sense that existing sub-populations may disappear but new ones cannot be created.

The model's dynamic is described by standard 'replication' equations. The replication dynamic is widely used in evolutionary models, which assume that the most profitable strategies proliferate in the population at the expense of others. Heckathorn (1996) describes this dynamic well:

"Based on the resulting payoffs, the actors with the most successful strategies proliferate at the expense of the less successful. This process is then repeated, generation after generation, until the system either approaches stable equilibrium or cyclical variation." (p. 261)

[5] This equilibrium represents a dilemma because the outcome of the game is non-cooperation when each player *individually* prefers mutual cooperation. It is well known that the outcome of the game depends essentially on two assumptions concerning rationality (individualism and instrumentality) and on an assumption concerning the type of interaction (anonymous).

[6] Note that the discount factor is related to the probability that the game will continue for another round. See Gintis (2009), p. 202-203.

This dynamics is usually employed in biology to study the evolution of species on the basis of the relative *fitness*. However, in social sciences there is a different interpretation of such a selection process: it involves learning by observing and imitating the behaviour of others. In what follows, we adopt neither the biological analogy nor the memetic one (i.e. the extension of gene-based biological evolution to meme-based social evolution). Instead, we use the concept of 'expected utility' as an indicator of the success (not necessarily material) of a strategy: a success which, over time, is imitated by less successful strategies (those with less expected utility). The dynamic of the model can be represented by the replication equations:

$$\dot{p}_i = p_i(Y_i - Y) \qquad i = 1,\dots,N \qquad [1]$$

where *p* denotes the proportion of subjects for each subpopulation, Y the average payoff, and Y_i the average payoff for a subject belonging to the subpopulation *i*.

The dynamic is defined on the invariant simplex:

$$\Delta = \left\{ p \in \Re^N, \sum_{i=1}^{N} p_i = 1, p_i \geq 0 \right\}$$

We shall use this analytical structure to analyse the evolutionary process that arises in a situation where there are two pure strategies, C and D, and first two, then three, and finally four subpopulations.

We are well aware that if the game is repeated, the possible strategies are infinite. We consequently restrict our analysis to four strategies, which we shall call (following Sugden 2004) B (= Brave) and C (= Cautious).

The strategies considered are therefore the following:

1. N: never cooperate. N is a highly important strategy because analysis of cooperation dynamics becomes non-banal precisely when non-cooperation scenarios are possible.

2. G: always cooperate.

3. B: a trigger strategy, which prescribes: cooperate with a player who has never not cooperated, do not cooperate with any other player who, in some previous round, has not cooperated, and *begin by cooperating*. In other words a player using B strategy will cooperate, but as soon as the opponent defects, the player using B strategy will defect for the reminder of the game. B stands for 'Brave', in fact. Bs are players who begin by cooperating (and therefore risk being 'exploited' by Ns or Cs in the first round). But if in the second round they do not receive cooperation, nor will they cooperate.

4. C: this trigger strategy has the same structure as B, the only difference being that C begins by not cooperating. The description is then: cooperate with a player who cooperated in some previous round and who has never not cooperated, do not cooperate with any other player and begin by not cooperating. If these cautious types are to cooperate, they must have obtained cooperation in the previous round. When Cs encounter other Cs or Ns, they never cooperate. An immediate consequence ensues: in a world with only Cs and Ns, cooperation will never be possible, and it will not be possible to distinguish Cs from Ns because they behave in exactly the same way.

If we use p_n, p_b, p_g, p_c to denote the probabilities of encountering, respectively, an N, B, G or C type, the expected utilities in a world with these four possible strategies are:

$$U_n = p_n(0) + p_b\beta + p_g \frac{\beta}{1-\pi} + p_c(0) \qquad [2]$$

An N type will never cooperate with other N types and with C types who begin by not cooperating and do not cooperate if the other player did not cooperate in the first round, whence $p_n(0)$, $p_c(0)$. If the N type encounters a B type, s/he will obtain β in the first round because B began with an act of cooperation, but the subsequent payoffs will be equal to 0 because B will stop cooperating from the second round onwards. Finally, if N encounters a G, s/he will obtain β in every round[7] because G will always cooperate.

$$U_b = p_n(-\gamma) + p_b \frac{(\beta-\gamma)}{1-\pi} + p_g \frac{(\beta-\gamma)}{1-\pi} + p_c\left(-\gamma + \beta\pi\right) \qquad [3]$$

The B type begins with an act of cooperation and continues to cooperate if the adversary in the first round has responded by cooperating. Cooperation is assured with other B types and with G types, but not with N types, or even with C types.[8]

$$U_g = p_n \frac{-\gamma}{1-\pi} + p_b \frac{(\beta-\gamma)}{1-\pi} + p_g \frac{(\beta-\gamma)}{1-\pi} - p_c\left(\frac{\beta-\gamma}{1-\pi} - \beta\right) \qquad [4]$$

A G type will therefore always cooperate with Bs and with Gs, and with Cs from the second round onwards, while Gs will let themselves be 'exploited' by Ns.

$$U_c = p_n(0) + p_b\left(\beta - \gamma\pi\right) + p_g\left(\frac{\beta-\gamma}{1-\pi} + \gamma\right) + p_c(0) \qquad [5]$$

Finally, a C type will not cooperate with Ns and Cs, and s/he will cooperate with Gs from the second round onwards. With Bs, C types will receive β in the first round, given that Bs begins with an act of cooperation, and (- γ) in the second round. From the third round onwards Cs will obtain 0.

3. The evolutionary analysis

In order to analyse the evolution in dynamic terms, we consider three strategies at a time (so that we can use simplexes).

After the first game, it is likely that the proportion of players adopting the winning strategy will increase in future pairings: that is, the winning strategy

[7] The expected utility associated to this interaction is hence $\beta + \beta\pi + \beta\pi^2 + \dots$, and then $\frac{\beta}{1-\pi}$.

[8] The payoff $p_c\left(-\gamma + \beta\pi\right)$depends on the fact that B cooperates the first time and C responds by not cooperating; B will therefore have (- γ), but C will cooperate in the second round, because B has cooperated in the first. From the third round onwards the payoff will be 0.

will be imitated by others. This will be the basis for our *both repeated and evolutionary* analysis.

It will be assumed in the analysis that $\pi > \frac{\gamma}{\beta}$.[9]

3.1. First case: N, C, G

We begin the analysis with B types omitted.

The replication dynamic can be represented with the following simplex:

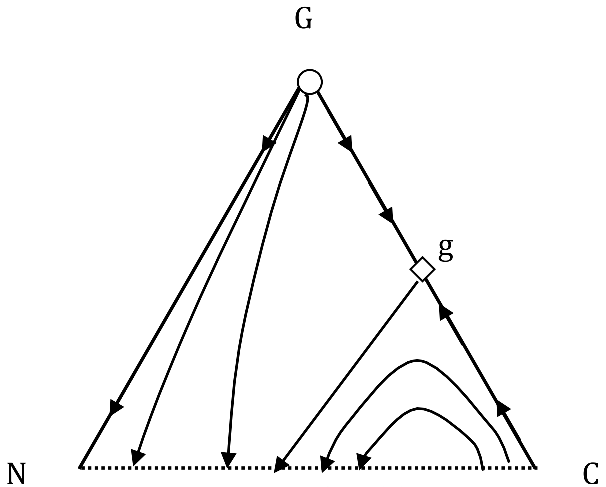


Figure 2.1: NCG

When strategies N, C and G are present, the outcome may be one of the multiple fixed points along the line NC, which signifies *non-cooperation*. If only strategies G and C are present, the outcome may be a unique combination of C and G that depends on the position of the point (in this case a saddle point), *g*, i.e.:

$$g \equiv \left(0, \quad \frac{\gamma(1-\pi)}{(\beta-\gamma)\pi} \quad \frac{\beta\pi-\gamma}{(\beta-\gamma)\pi} \right).$$

[9] We imagine, in fact, that two players agree to cooperate in each round. If they abide by the agreement, the expected utility of each player is $\frac{\beta-\gamma}{1-\pi}$. However, if one of the players breaks the agreement, the other will no longer cooperate. Thus a player who breaks the agreement in the first round will receive β, but from the second round onwards s(/he will always receive 0. The condition for cooperation agreements (without enforcement) to come about is: $\frac{\beta-\gamma}{1-\pi} > \beta$, and hence $\pi > \frac{\gamma}{\beta}$. On this see also Sugden (2004).

Other conditions remaining equal, if $\pi \to 1$ – so that the likelihood of continuing with the same person initially encountered is very high – the point shifts towards vertex G.

This result strikes us as important: only G types are able somehow to activate Cs, who without Gs would always be confined to a world of non-cooperation. The following proposition therefore holds:

Proposition 1. In a world in which the types or strategies N, G, C are present, the replication dynamic has two different outcomes: a combination of C and G (fixed point g) only if p_n is equal to 0, or a combination along the line of fixed points N, C (and consequently non-cooperation).

Without the presence of B types – who always begin with an act of cooperation – it is unlikely that virtuous cooperation mechanisms will be triggered.

3.2. Second case: N, B, C

Another interesting case is that in which G types are absent. Here too, non-cooperation is a probable equilibrium. The other equilibrium is the one where only B strategies survive. In a three-strategy world in which only Ns, Cs and Bs are present, in fact, Ns and Cs will never cooperate, and moreover the Ns will have no Gs to exploit. Instead, the Bs will cooperate only and exclusively with each other, obtaining a greater payoff – if the game lasts for a long time – than that received by the Ns and the Cs.

Here too, as shown by figure 2.2, the possible long-period equilibrium depends on the coordinates of the fixed point *f*.

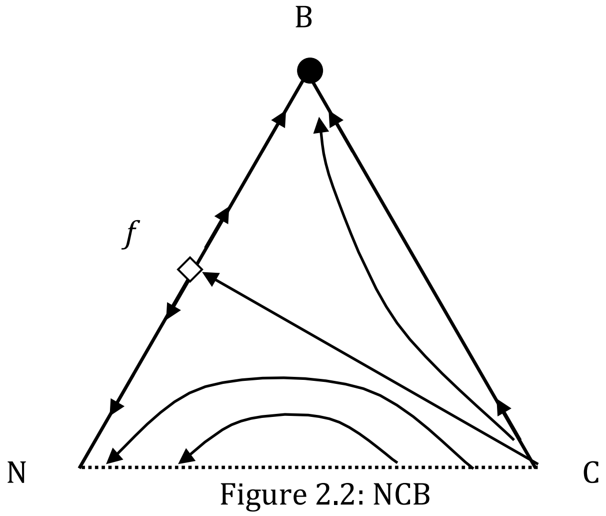


Figure 2.2: NCB

All the points of departure in the simplex lying below the trajectory from C to *f* will evolve towards a non-cooperative equilibrium if N and C are present.

Proposition 2 . *In a world in which the types or strategies N, C, B are present, the replication dynamic has two different outcomes: the survival of B strategies alone, or a combination along the line of fixed points N, C (and consequently non-cooperation).*

The coordinates of point *f* are now:

$$f \equiv \begin{pmatrix} \frac{\beta\pi - \gamma}{(\beta - \gamma)\pi} & 0 & \frac{\gamma(1-\pi)}{(\beta-\gamma)\pi} \end{pmatrix}$$

It is evident that if $\pi \to 1$, the point tends to shift towards the N vertex, so that that greater the probability of the game continuing, the more likely it becomes that Bs will prevail and that the cooperative outcome will occur. In a world without G types, Cs do not begin to cooperate. We may say that the sacrifice of the Gs somehow restores cooperation potential to Cs, for without their presence the only possible form of cooperation is that between B types. To be noted is that B types begin with an act of cooperation. In their absence, a non-cooperative equilibrium would arise.

3.3. Third case: N, B, G

The simplex relative to this third and final case shows that, depending on the point of departure and the position of fixed point f on the side NB, there will be a different final equilibrium, which may be a combination of G and B, or a world consisting only of Ns. Matters are different when the three types instead coexist in the population at time 1 (when the dynamic begins). In this case, non-reciprocity, i.e. an equilibrium consisting of only N types, may prevail.

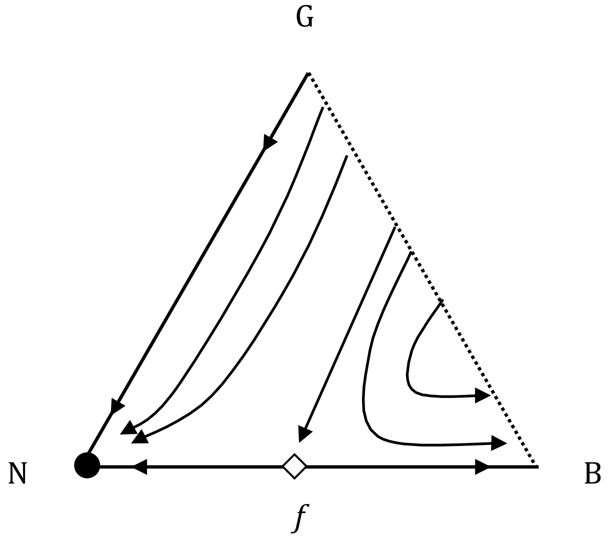


Figure 2.3: NBG

Proposition 3. In a world in which the strategies N, B, G are present, two equilibria are possible: the survival of only types N and a coexistence of B types and G types along the line of fixed points on the B-G side. Which of the two equilibria will come about depends on the position of the fixed point f along the N-B side.

As the simplex is constructed here, considering that the position of N in terms of fraction of the population is (1,0,0) and the position of B is (0,1,0), the fixed point *f* has the following coordinates:

$$f \equiv \left(\frac{\beta\pi - \gamma}{(\beta-\gamma)\pi},\ \ \frac{\gamma(1-\pi)}{(\beta-\gamma)\pi},\ \ 0 \right)$$

The position of point *f* therefore depends on β and γ, and on the value of π. In particular, for $\pi \rightarrow \frac{\gamma}{\beta}$, point *f* will approach B. If instead $\pi \rightarrow 1$, point *f* will shift towards N. With a small value of π, *ceteris paribus*, the likelihood that only N types will prevail is very high; instead, with a very high π, it very likely that the final equilibrium will be the one in which B types and G types coexist.

For every intermediate value between the two extremes, the final equilibrium will depend on the point of departure: if this is a point to the left of the trajectory leading from side B-G to point *f*, then the tendency is an equilibrium of only Ns; vice versa, if the point of departure is to the right of the trajectory, the outcome will be a coexistence of Bs and Gs. Note that points to the left are characterized, amongst other things, by a lower percentage of Bs than of Gs. It is therefore important that B types be relatively more than Gs and Ns for the B-G equilibrium to come about. In short, evident here is the delicate role of G strategies: if there are too many of them, they foster the emergence of N types over Bs. Metaphors aside, in a population where non-cooperation is possible, if there are too many unconditional acts, not only are they likely to become extinct, but they will also extinguish the possibility of cooperation, for an equilibrium consisting of non-generalized cooperation.

At the same time, the coordinates of point *f* also depend on *β* and *γ*. The value of *γ* is the one which most clearly tells us what the social rewards structure is. A high *γ* denotes a culture which penalizes reciprocity, while a high (*β*-*γ*) denotes a culture which rewards it. In fact, if the first coordinate is high, point *f* tends to N (the same happens if the second coordinate is low), while if it is low *f* tends to B.

This is because the coordinate of N is directly proportional to *β*: while both coordinates depend on (*β*-*γ*), the sign of *γ* is negative in the coordinate of N and positive in the coordinate of B. This tells us that the more a society, *ceteris paribus*, makes reciprocity of G and B type costly, the more likely the prevalence of non-cooperation becomes.

4. In a four-dimensional world

Thus far we have compared three strategies at a time, and we have analysed their dynamic evolution. The question now is what changes if the four strategies N, B, G, C interact simultaneously.

In the four-strategies case, the replication dynamic can be depicted by a three-dimensional simplex.:

$$\Delta = \left\{ \mathrm{p} \in \Re^4 : \mathrm{p} \geq 0 \ \text{ e } \ p_n + p_b + p_g + p_c = 1 \right\}$$

In this case matrix A becomes:

$$A = \begin{pmatrix} 0 & \beta & \frac{\beta}{1-\pi} & 0 \\ -\gamma & \frac{\beta-\gamma}{1-\pi} & \frac{\beta-\gamma}{1-\pi} & \beta\pi-\gamma \\ \frac{-\gamma}{1-\pi} & \frac{\beta-\gamma}{1-\pi} & \frac{\beta-\gamma}{1-\pi} & \frac{\beta\pi-\gamma}{1-\pi} \\ 0 & \beta-\pi\gamma & \frac{\beta-\pi\gamma}{1-\pi} & 0 \end{pmatrix} \qquad [7]$$

The system of equations becomes:

$$\begin{aligned} \dot{p}_n &= p_n\left[(Ap)_1 - {}^t\mathrm{p}\cdot\mathrm{Ap}\right] \\ \dot{p}_b &= p_b\left[(Ap)_2 - {}^t\mathrm{p}\cdot\mathrm{Ap}\right] \\ \dot{p}_g &= p_g\left[(Ap)_3 - {}^t\mathrm{p}\cdot\mathrm{Ap}\right] \\ \dot{p}_c &= p_c\left[(Ap)_4 - {}^t\mathrm{p}\cdot\mathrm{Ap}\right] \end{aligned} \qquad [8]$$

where
${}^t\mathrm{p} \equiv (p_n, p_b, p_g, p_c)$

Given that analysis of the system of differential equations [8] would be highly complex, here we only report the frontier conditions (those in which at least one strategy is extinct). Following the examples of Hirshleifer and Martinez Coll (1991), and of Antoci, Sacco and Zarri (2004), we may represent the surface (or frontier) of Δ on the plane. The simplex Δ can be imagined as having a triangular base N,C,B, and G as its upper vertex (if the simplex in figure 2.4 were drawn three-dimensionally, the three vertices G would become a single upper vertex).

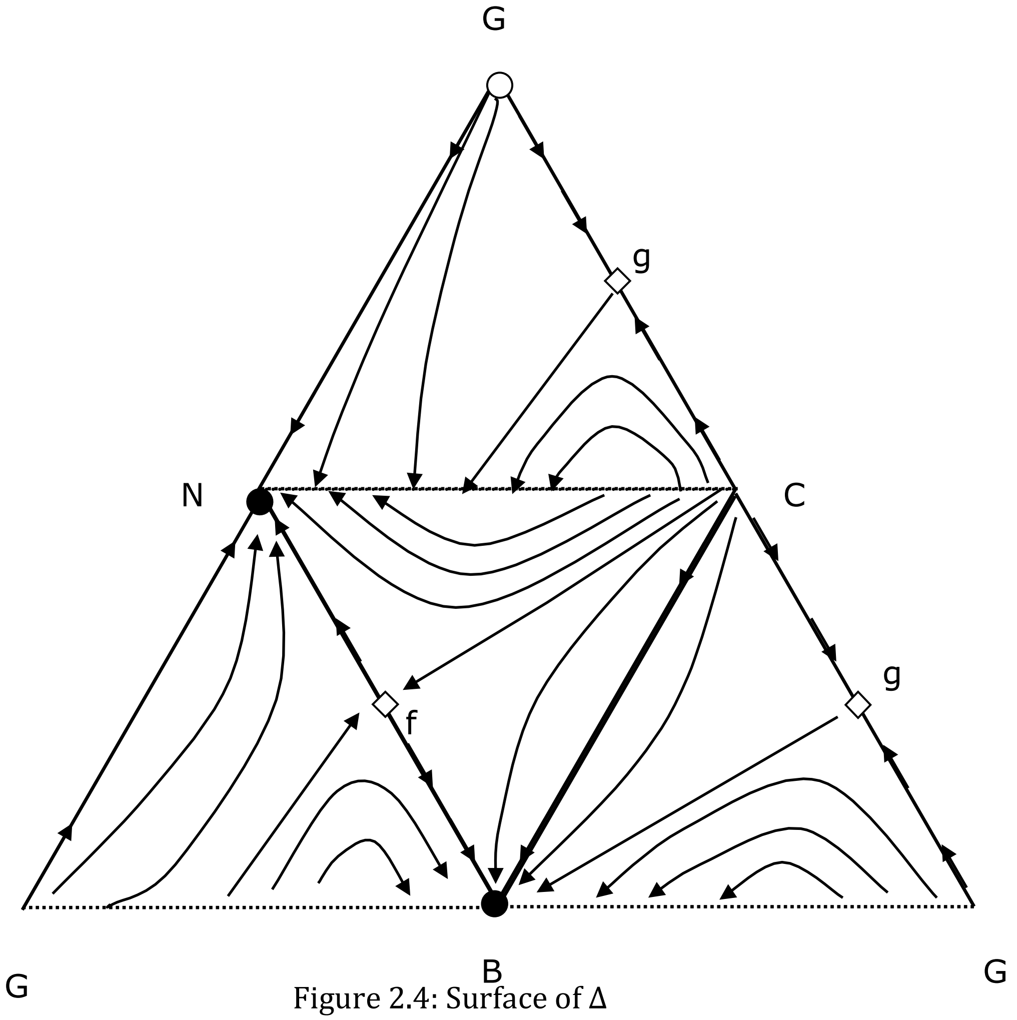


Figure 2.4: Surface of Δ

Figure 2.4 shows that there are four possible equilibrium combinations:

- a combination of G and B, i.e. cooperation
- a combination of N and C, i.e. non-cooperation
- the extinction of all the strategies except N
- the extinction of all the strategies except B.

4.1. Some simulations

Which of these equilibria are more likely depends on the initial conditions. To furnish a clearer idea of the dynamic, we now report some simulations. They have been obtained by setting various initial conditions for the system. We assigned the following values to the parameters:

$\beta = 2, \gamma = 1, \pi = 4/5$

The first graph shows the evolution over time of the strategies when the initial conditions state: $p_n = p_b = p_g = p_c = 0.25$.

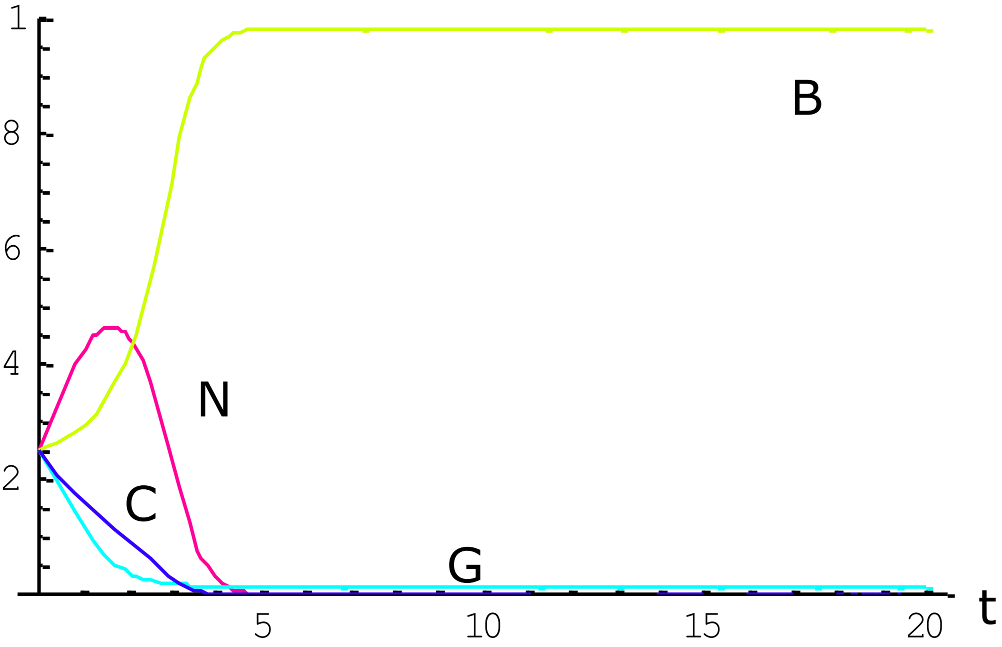


Figure 2.5: evolution over time with $p_n = p_b = p_g = p_c = 0.25$.

In this case the final equilibrium is of the B-G type where the proportion of G is very small. What happens if we change the initial conditions? The next graph illustrates a situation where the initial proportions are $p_n = 0.25$, $p_b = 0.25$, $p_g = 0.1$ $p_c = 0.4$. We have left the proportions of B and G unaltered, but we have increased Cs with respect to Gs.

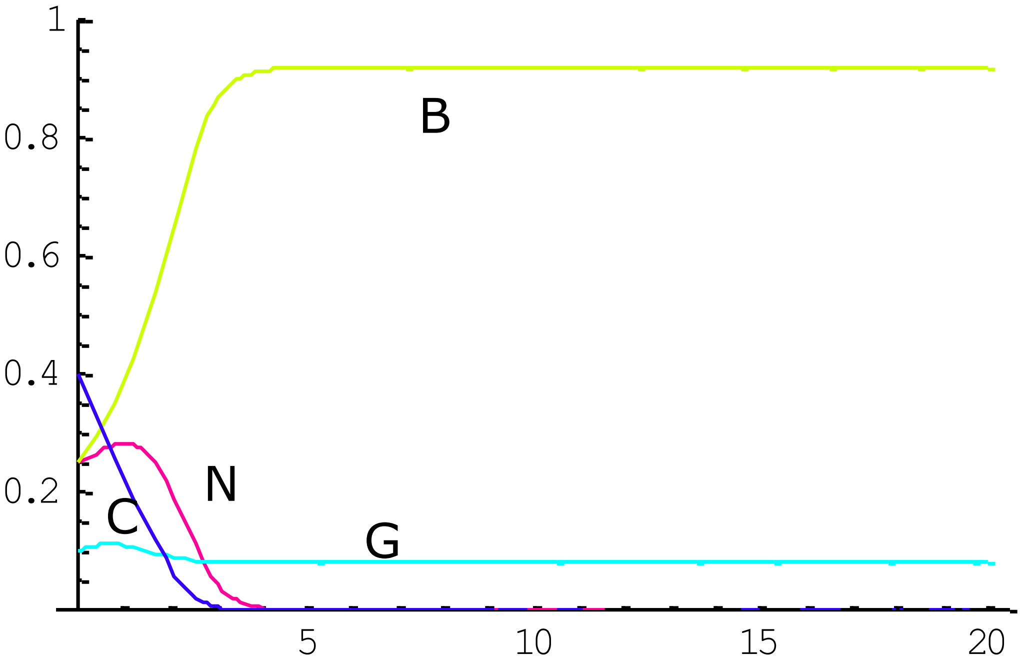


Figure 2.6: evolution over time with p_n = 0.25, p_b = 0.25p_g = 0.10 p_c = 0.4

Interestingly, a greater proportion of Cs, although it does not improve their chances of 'survival', helps the development of Gs, which in this case remain constant over time. We saw in section 3.1.1 that only G types are able to activate Cs; we may now state that Cs are essential for the survival of Gs. The importance of the role performed by Cs (which in the three-strategy world seemed almost irrelevant) also emerges from the following graph, which has been constructed with the following initial proportions: p_n = 0.4, p_b = 0.3, p_g = 0.1 p_c = 0.2. In this case the Ns are initially in a greater proportion than Bs, and there are more Cs than Gs.

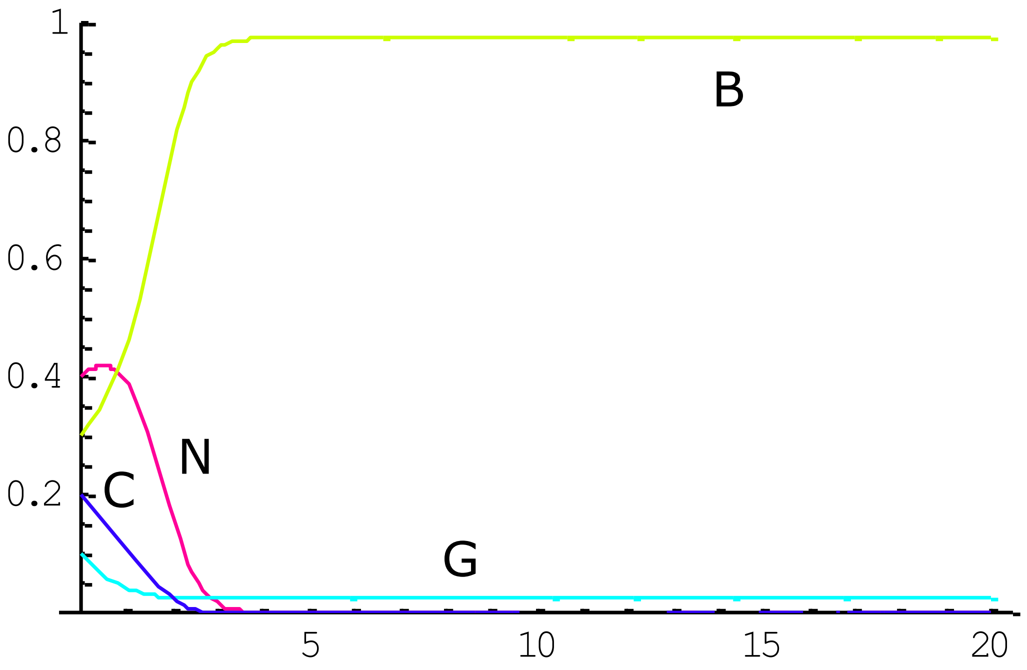


Figure 2.7: evolution over time with $p_n = 0.4$, $p_b = 0.3$, $p_g = 0.10$ $p_c = 0.20$

Hence, cooperation may prevail even if there are initially more Ns than Bs, provided that there is a sufficient number of Cs.

5. conclusions

Now we may draw some conclusions, that can be summarised as follows.

(a) *The 'crucial' role of G types*. We have seen at various points in our analysis that G types should not be too numerous, because if they are they compromise themselves and also the survival, for example, of Bs. In populations where non-cooperation is possible (which is the case of all real ones), unconditional acts are essential, but when too numerous, they become counter-productive.

(b) G types perform a vital role, for only they can activate the cooperation of Cs. Without the presence of G types, Cs would never experience cooperation and therefore would never respond with an act of cooperation. G types are consequently valuable, but they should be protected. The success of numerous forms of cooperation – from firms to families – depends also, and sometimes above all, on the presence of a small number of unconditional reciprocators able to activate people who would never be so activated if they only interacted with conditional cooperators.

(c) *Alliances: C types*. These are 'activated' by Gs, but at the same time their presence is highly beneficial to Gs because it increases their expected utility. Gs, in fact, cooperate with Bs and with Cs, but they are exploited by Ns. In a four-strategy world, Cs protect the Gs against extinction.

Cooperation is therefore favoured by heterogeneity.

From a mathematical point of view, it might be objected that G types are not necessary. The onset of cooperation would only require slightly more sophisticated Bs[10]. But this was not the purpose (i.e. to study which strategies favour cooperation) for which the model was conceived. Our analysis started from the assumption that behaviours like G exist in civil society. (And who could deny the presence in the real world of unconditional actions? Even Binmore (2006) with his orthodoxy and anthropological parsimony admits their existence). Our model has sought to analysis the conditions under which unconditional actions can not only survive but also perform a virtuous civil role.

[10] Note that if we use Tit-fot-tat strategies instead of trigger strategies for B and C the results do not change. The expected utility for B when he meets C will be $\frac{-\gamma+\beta\pi}{1-\pi^2}$ instead of $-\gamma + \beta\pi$ and the expected utility for C when he meets B will be $\frac{\beta-\gamma\pi}{1-\pi^2}$ instead of $\beta - \gamma\pi$. This does not affect the resulting equilibria.

Appendix 1. Mathematical proofs

Proof of proposition 1:

The expected utilities are:

$$U_n = p_n(0) + p_c(0) + p_g\left(\frac{\beta}{1-\pi}\right)$$
$$U_c = p_n(0) + p_c(0) + p_g\left(\frac{\beta-\gamma\pi}{1-\pi}\right)$$
$$U_g = p_n\left(\frac{-\gamma}{1-\pi}\right) + p_c\left(\frac{\beta\pi-\gamma}{1-\pi}\right) + p_g\left(\frac{\beta-\gamma}{1-\pi}\right)$$

The matrix of payoffs is:

$$\begin{pmatrix} 0 & 0 & \frac{\beta}{1-\pi} \\ 0 & 0 & \frac{\beta-\pi\gamma}{1-\pi} \\ \frac{-\gamma}{1-\pi} & \frac{\beta\pi-\gamma}{1-\pi} & \frac{\beta-\gamma}{1-\pi} \end{pmatrix}$$

Adding a constant to each column of A does not change the dynamics, so we subtract the first row:

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{-\pi\gamma}{1-\pi} \\ \frac{-\gamma}{1-\pi} & \frac{\beta\pi-\gamma}{1-\pi} & \frac{-\gamma}{1-\pi} \end{pmatrix}$$

We know that:
$\beta > \gamma > 0$

Following Bomze (1983), proposition 1 (p. 210) :

1. the eigenvalue of the corner N in direction N-C is equal to 0
2. the eigenvalue of the corner N in direction N-G is proportional to $\frac{-\gamma}{1-\pi}$, and then is negative
3. the eigenvalue of the corner C in direction C-N is equal to 0
4. the eigenvalue of the corner C in direction C-G is proportional to $\frac{\beta\pi-\gamma}{1-\pi}$, then is positive (we have supposed that $\pi > \frac{\gamma}{\beta}$)
5. the eigenvalue of the corner G in direction G-C is proportional to $\frac{\gamma-\gamma\pi}{1-\pi}$ and then is positive

6. the eigenvalue of the corner G in direction G-N is proportional to $\frac{-\gamma}{1-\pi}$, and then is positive.

Following proposition 2 (p. 210) we know that N-C is pointwise fixed.

Proposition 5 (p. 211) tells us that there exists a fixed point g (saddle point) on the side G-C, in fact the quantity *(e – b)(f – c)* is negative, and the eingenvalues associated to the fixed point are proportional to:

1.$\frac{(e-b)(c-f)}{e-b+c-f}$, that means $\frac{-\left(\frac{\beta\pi-\gamma}{1-\pi}\right)\left(\frac{\gamma-\pi\gamma}{1-\pi}\right)}{\left(\frac{\beta\pi-\pi\gamma}{1-\pi}\right)}$: this quantity is negative;

2.$\frac{bf-ce}{e-b+c-f}$, that is positive.

Proof of proposition 2:

Expected utilities:

$U_n = p_n(0) + p_c(0) + p_b(\beta)$

$U_c = p_n(0) + p_c(0) + p_b(\beta - \gamma\pi)$

$U_b = p_n(0) + p_c(0) + p_b\left(\frac{\beta - \gamma}{1 - \pi}\right)$

Matrices:

$$\begin{pmatrix} 0 & 0 & \beta \\ 0 & 0 & \beta - \gamma\pi \\ -\gamma & \beta\pi - \gamma & \frac{\beta-\gamma}{1-\pi} \end{pmatrix} \text{and} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\pi\gamma \\ -\gamma & \beta\pi - \gamma & \frac{\beta\pi-\gamma}{1-\pi} \end{pmatrix}$$

1. the eigenvalue of the corner N in direction N-C is equal to 0
2. the eigenvalue of the corner N in direction N-B is proportional to $-\gamma$, and then is negative
3. the eigenvalue of the corner C in direction C-N is equal to 0
4. the eigenvalue of the corner C in direction C-G is proportional to $\beta\pi - \gamma$, and then is positive
5. the eigenvalue of the corner B in direction B-C is proportional to $\frac{-\beta\pi+\gamma-\pi\gamma+\pi^2\gamma}{1-\pi}$ and then is negative
6. the eigenvalue of the corner B in direction B-N is proportional to $\frac{\gamma-\beta\pi}{1-\pi}$, and then is negative

Following proposition 2 (p. 210) we may say:

- the side N-C is pointwise fixed
- On the side N-B there exists an unique fixed point *f*; the eigenvalues of f are positively proportional to:

γ (positive)

$\frac{0-(-\gamma\pi)(-\gamma)}{\frac{\beta\pi-\gamma}{1-\pi}}$ (negative).

The fixed point has coordinates(Bomze 1983, p. 204):

$$p_n = \frac{1}{1+\frac{\gamma}{\frac{\beta\pi-\gamma}{1-\pi}}} = \frac{\beta\pi-\gamma}{(\beta-\gamma)\pi}$$

$$p_c = 0$$

$$p_b = \frac{\gamma(1-\pi)}{(\beta-\gamma)\pi}$$

We know that do not exist fixed points on the side C-B (prop. 5) and that do not exist internal fixed points (prop.6).

Proof of proposition 3:
Expected utilities:

$$U_n = p_n(0) + p_b(\beta) + p_g\left(\frac{\beta}{1-\pi}\right)$$
$$U_b = p_n(-\gamma) + p_b\left(\frac{\beta-\gamma}{1-\pi}\right) + p_g\left(\frac{\beta-\gamma}{1-\pi}\right)$$
$$U_g = p_n\left(\frac{-\gamma}{1-\pi}\right) + p_b\left(\frac{\beta-\gamma}{1-\pi}\right) + p_g\left(\frac{\beta-\gamma}{1-\pi}\right)$$

Matrices:

$$\begin{pmatrix} 0 & \beta & \frac{\beta}{1-\pi} \\ -\gamma & \frac{\beta-\gamma}{1-\pi} & \frac{\beta-\gamma}{1-\pi} \\ \frac{-\gamma}{1-\pi} & \frac{\beta-\gamma}{1-\pi} & \frac{\beta-\gamma}{1-\pi} \end{pmatrix}, \text{and} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{\beta\pi-\gamma}{1-\pi} & \frac{-\gamma}{1-\pi} \\ \frac{-\gamma}{1-\pi} & \frac{\beta-\gamma}{1-\pi} & \frac{-\gamma}{1-\pi} \end{pmatrix}$$

1. the eigenvalue of the corner N in direction N-B is proportional to $-\gamma$ and then is negative
2. the eigenvalue of the corner N in direction N-G is proportional to $\frac{-\gamma}{1-\pi}$ and then is negative
3. the eigenvalue of the corner B in direction B-N is proportional to $-\frac{\beta\pi-\gamma}{1-\pi}$ and then is negative
4. the eigenvalue of the corner B in direction B-G is equal to zero

5. the eigenvalue of the corner G in direction G-B is equal to zero
6. the eigenvalue of the corner G in direction G-N is equal to $\frac{\gamma}{1-\pi}$, and then is positive

We know that there esists a fixed point on the side N-B (prop.2) , and that the eigenvalues of the fixed point are positively proportional to :

γ, then positive
$\frac{\left(\frac{\beta\pi-\gamma}{1-\pi}\right)\left(\frac{-\gamma}{1-\pi}\right)-(-\gamma)\left(\frac{\beta\pi-\gamma}{1-\pi}\right)}{\frac{\beta\pi-\gamma}{1-\pi}}$ that becomes:$\frac{-\gamma\pi}{1-\pi}$ and then is negative.
The fixed point has coordinates (Bomze 1983, p. 204):

$$p_n = \frac{1}{1+\frac{\gamma}{\frac{\beta\pi-\gamma}{1-\pi}}} = \frac{\beta\pi-\gamma}{(\beta-\gamma)\pi}$$

$$p_b = \frac{\gamma(1-\pi)}{(\beta-\gamma)\pi}$$

$$p_g = 0$$

We also know that the side B-G is pointwise fixed.

Chapter 3:
We-thinking and vacillation between frames: *Filling a gap in Bacharach's theory*

1. Introduction

The idea of team-thinking or we-thinking is increasingly drawing the attention of scholars. In its general formulation, it has 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[11]) have developed analytical frameworks from an economic point of view. We-thinking theories allow groups to deliberate as agents. A central concept in these theories is what has been called *team reasoning*:

> "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 we-reason they seek an answer to the question: "What should we do?", and they act accordingly.

The main claim of scholars who analyze we-thinking is that it is a coherent mode of reasoning people may use when they face a decision problem of a certain type.

[11] The 2006 book was published after Bacharach's death. The editors, Natalie Gold and Robert Sugden, assembled all the existent materials Bacharach intended to put into the book and added their own discussion of Bacharach's plans for the chapters that were uncompleted when he died.

We-thinking theories have been introduced into the economic domain for both theoretical and empirical reasons.

First of all, we-thinking theories account for the relational nature of humankind (see Sugden 2005, Bruni 2008, and Davis 2009). As Hollis puts it: "we need a more social conception of what persons are and a role-related account of the obligations which make the social world go round and express our humanity" (Hollis 1998, p. 104).

Secondly, team reasoning helps to solve some puzzles that arise in game theory, especially linked to Hi-Lo[12] and one-shot Prisoner's Dilemma (PD) games, in which rational choice theory cannot explain selection of the Pareto-superior equilibria or cooperation[13].

We thinking is also a way to explain how people can coordinate on 'focal point' equilibria: focal points have been introduced by Schelling (1960), they are particular Nash Equilibria on which the players' expectations converge. Team reasoning offers an explanation of coordination on focal points, which has been tested by Bacharach and Bernasconi (1997), Barsdley et al. (2010), and Crawford et al. (2008).

Finally, team reasoning can also explain experimental and empirical evidence on how people behave in other games and decision contexts. Experimental evidence shows that, especially in some kinds of games, people do endorse we-thinking[14]. In particular Colman et. al. (2008), making use of likelife vignettes and abstract games, show evidence for team reasoning, as a good predictor of strategy choices.

Nevertheless there are different opinions about the way in which we-thinking arises and how it brings people to behave in a particular way.

[12] The Hi-Lo game is in general a *n* player game in which each player chooses one item from the same set of alternatives. Each alternative is associated with a prize, and one alternative's prize is greater than the others. If all players choose the same alternative they get the associated prize, if not nobody gets anything.

[13] Bacharach 2006 (pp. 44-58) refers to the most relevant models that have faced this issue, explaining why a new theory is needed; Colman (2008) explains how team reasoning provides a justification for choosing payoff-dominant equilibria, a concept introduced by Harsanyi and Selten (1988).

[14] See Guala et al. (2009), Becchetti, Faillo, Degli Antoni (2009), and Tan and Zizzo (2008) for a review of experiments.

Different authors have proposed different conceptual analyses of the issue, with no general agreement among them[15].

Among the few scholars who have proposed formal approaches to illustrating how we-thinking arises, Bacharach offered one of the most developed theories from the game theoretic point of view[16].

He proposes a formal theory of games with I-reasoners and we-reasoners, with the mode of reasoning taken as given. A fundamental point in Bacharach's theory is that the determination of mode of reasoning is a psychological matter, prior to rational choice, and is given by frames. So, as he recognizes, to complete the theory he needs to build a theory which explains which mode of reasoning will be in play. This means to endogenize I/we determination. This part of Bacharch's theory is less developed, although he suggests some intuitions, not always mutually consistent and not fully developed. He tries to complete his theory following two different approaches: the concept of the *harmony* of the game, which has been further developed by Tan and Zizzo[17] , and the *interdependence hypothesis,* which links to an underdeveloped intuition about vacillation between frames. Because of his premature death, which occurred unexpectedly, he never achieved his aim of endogenising the determination of the mode of reasoning.

In the present paper, I shall suggest a way to complete Bacharach's theory, generalising the interdepence hypothesis and building on his intuition about vacillation. I propose a formal model of vacillation between frames, which allows individuals to switch from I to we mode of reasoning and vice versa (section 4 and 5). It is a simple model based on an intuition that Bacharach expressed but did not fully develop, concerning double-crossing in the PD game. In order to develop my proposal, Bacharach's theory of team-

[15] See the chapter 1 for a review of team reasoning theories.

[16] Of course Bacharach is not the only theorist who has adressed team reasoning from a game-theoretic point of view: among theorists who use team reasoning or related principles, such as 'principle of coordination', for explaining focal points, Sugden (1995), Casajus (2000), and Janssen (2001, 2006) make use of game theory; Zizzo (2004) and Zizzo and Tan (2003) do so introducing a 'game harmony' measure. Nevertheless, Bacharach is the only one who tries to explain the emergence of team reasoning using game theory.

[17] See Zizzo and Tan (2007), Tan and Zizzo (2008).

reasoning will be analysed in section 2, by taking into account published and unpublished material. In section 3 I propose a discussion of some not fully developed intuitions of Bacharach, and section 6 presents the conclusions.

2. Bacharach's theory of we-thinking

> "The answers to fundamental questions about coordination and cooperation… lie in the agent's conception not of the objects of choice, nor of the consequences, but of herself and of the agents with whom she is interacting" (Bacharach 2006, p. 70).

This sentence is the starting point of Bacharach's analysis of we-reasoning. We-reasoning is seen as a powerful 'mechanism' (in Bacharach's words) for solving puzzles about coordination and cooperation in game theory (i.e. games like Hi-Lo and PD). In his work Bacharach tries to demonstrate that we-reasoning is a valid mode of reasoning and that people do endorse it[18].

Bacharach's main purpose is to explain cooperation, seen as a successful group activity (ib p. 69), and the core mechanism for doing that comprehends 'framing', 'common purpose', and 'cooperation':

> "(i) we frame ourselves as members of groups; (ii) …perceived agreement of individual goals among a set of individuals favours framing as members of a group with this common goal; (iii) the group framing tends to issue in efficient cooperation for the group goal" (ib p. 90).

In what follows, I illustrate the building blocks of Bacharach's theory, but, first of all, I give an account of how and when Bacharach developed the idea of we-thinking. This is because the particular pattern he followed could

[18] Bacharach claims that there are five kinds of evidence in support of this claim: logical, introspective, evolutionary, transcendental and experimental. In particular he gives an account of two experiments, one conducted by himself and Guerra, and the other by himself and Bernasconi, which provide some behavioural evidence that group identification leads people to we-reasoning (see Bacharach 2006, pp.145-146, and Bacharach and Bernasconi 1997).

offer hints for developing some of his intuitions, remaining faithful to his thought.

2.1. Development of Bacharach's thought

Bacharach started by building Variable Frame Theory (Bacharach 1993), when in parallel he was developing a theory of cooperation. Variable Frame Theory (VFT) is an analysis of choices in games in which frames are taken into account. VFT allows games with descriptions of players' frames. Concisely, in VFT a player can intentionally choose an object, or an action, only if she has a way of thinking about that object or that action, i.e. she has a frame in which it is represented. Frames can be more or less salient or available, depending on a probability measure on them.

Bacharach's aim in developing VFT was to explain the choice of focal points in games: by making use of VFT he could turn focal point problems into Hi-Lo games. We-thinking theory, as proposed by Sugden (1993), helped him to explain the selection of Pareto-superior equilibria in Hi-Lo and in coordination games more generally. He started then, to develop his own theory of we-reasoning.

In 1995 he introduced the category of 'fellow member reasoner':

> "Someone who is a member of a natural type T and chooses a certain strategy if she is sufficiently sure that her interactants are also member of T" (1995, p.1).

In this context he tries to link T-membership to VFT and, at the same time he introduces the 'we' category:

> "The present paper has made type T membership an issue which type T members think about, and nuanced their capacity to recognize it. An alternative development would make T membership a variable element in players' frames in the sense of variable frame theory: that is, a player might or might not think about the game in terms of whether she and her coplayers belong to T. In the case in which T is the player

> set, we may put this by saying that a player may or may not think in 'we' terms about how to play the game. The more inclined a player is to 'we' thinking, and the more inclined she takes coplayers to be, the more will fellow-member reasoning be favoured" (p.17).

In 1997 Bacharach formally introduces we-thinking, in an unpublished paper whose title is: "'We' Equilibria: A Variable Frame Theory of Cooperation". The first published paper in which Bacharach formalizes his theory is an article published in 1999 about 'interactive team reasoning'. In it Bacharach introduces some elements that we can find in the book, such as group identification, team reasoning as the effect of group identification, and unreliable team interaction (which in the book becomes cirscumspect team reasoning). Between the 1999 article and the book we may find some lecture notes, in which the concepts of agency and 'superagency' begin to appear. The book represents an (incomplete, because of his death) attempt to build a complete, and at the same time simple, theory of we-thinking: I shall present in the following subsections the theory as it appears in the book.

2.2. I-reasoning and we-reasoning

First of all, Bacharach allows for the existence of both I and we modes of reasoning. Each is seen as rational maximization of a von Neumann - Morgenstern utility function. I-reasoning is represented by a standard utility function. We-reasoning, instead, requires a team utility function (W): "a game-theoretic treatment of agents who may group-identify must... determine a payoff function to represent the group objective" (Bacharach 2006 p. 87)[19]. In order to clarify what the group identification process implies about what the players want as a group, or, in other words, in order to clarify what W is, Bacharach proposes that W must satisfy the 'Paretianness' condition:

[19] In a previous article Bacharach said: "Our theory takes the line that collections of people can have objectives" (1999, p. 120).

if a profile of actions p is weakly Pareto-superior to p'

then $W(p) \geq W(p')$.

This requires that group objectives are related to personal ones. Examples of group utility functions include the utilitarian function and weighted utilitarian functions. According to Bacharach, the most commonly used form of group utility function is the mean of the individual payoff[20]. Another important point for Bacharach is to allow principles of symmetry and fairness between individual payoffs[21] to be embedded in W.

2.3. Frames

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).

In Bacharach's framework a person may start to we-reason only if she has 'we' concepts in her frame. If the we-frame is active in a subject she begins to think of herself as a part of a collective actor, then she begins to team-think, and this means that in the face of a decision problem she will answer the question: "What shall we do?". In Bacharach's theory then, to see the we-frame implies to endorse that frame. In his theory group identification is a

[20] See Bacharach 1999, p. 126. In order to derive group utility functions from the individual ones, interpersonal comparisons are required. Bacharach is implicitly assuming that these are possible. Interpersonal comparisons do not affect the way in which players play, but they are required to construct the we-utility function. For a discussion of properties of W see Gold (2012).

[21] "Such as those of Nash's axiomatic bargaining theory"(Bacharach 2006, p. 88).

framing phenomenon that determines choices by "changing the logic by which people reason about what to do" (ib). When reasoning in the individual standard mode (I-reasoning), an agent looks at a decision problem by thinking what it would be best for her to do. Group identification changes this logic to we-reasoning: the agent will think: "What would best be for us to do?".

2.4. Circumspect team reasoning

One of Bacharach's aims is to explain situations in which some people may we-reason and some others may not. In order to model these situations, he assumes that the 'we' frame comes to mind with probability ω, which represents the probability that a subject group-identifies. The probability ω is common knowledge amongst team members[22]: "in coming to frame a situation as a problem 'for us', an individual also gains some sense of how likely it is that another individual would frame it in the same way" (ib p. 163). A context in which some people may group-identify and some may not is seen by Bacharach as an unreliable coordination context, and team reasoning in this context is called circumspect team reasoning. Briefly, people who we-reason in an unreliable coordination context look for the best available profile o - the combination of actions - that maximizes W given that each person will choose to do her part in o with probability ω, or will fail - i.e. act on I-reasoning - with probability (1- ω). One problem which remains open in Bacharach's theory is the endogenization of ω: he sees the need for endogenization and proposes some speculations, but he did not complete this part of the theory, as we shall see later.

2.5. Variable Frame Theory and 'vacillation'

Bacharch's (never reached) aim was to explain we-reasoning in terms of Variable Frame Theory (VFT).

[22] In a previous work (1999), Bacharach has developed a more formalized model, in which each agent can participate or lapse in a team and everyone, before choosing, receives a signal knowing the joint probability distribution of this signal and agent's state (i.e. an agent's signal includes her participation state).

The intersection between VFT and we-thinking would have been called by Bacharach 'Variable Agency Theory' (Bacharach 2006, p.59). However, the completion of the description of we-reasoning in terms of VFT raised problems that he had not solved at the time of his death. Let us see these problems.

In Bacharach's circumspect team reasoning, as I have said before, if people group identify, then the we-frame comes to their mind and they start to we-reason. It seems as though in Bacharach's framing theory there are two aspects that are deeply linked: in framing a situation, the first step is to recognize a frame, that is, coming to see it; the second step is endorsing that frame, that is, reasoning as the frame allows you to do. In Bacharach's theory group identification means not only coming to see a particular way of reasoning, but also endorsing it. The 'compression' between the two aspects of framing is due to VFT. However, in the original form of VFT, changing frame does not mean to change the way of reasoning, and the decision problem for a subject is fully determined by the interplay of his frame and the objective world. VFT was originally thought of as a way to allow of a player to frame different situations differently, but frames were not related to different agencies. In constructing his Variable Agency Theory, Bacharach was trying to use VFT in a new way, but, because of this 'compression', he could not allow people to use more than one frame at a time. In a certain sense, as it has been noticed by Gold and Sugden (in Bacharach 2006), in the we-frame people become committed to we-reasoning:

> "In the theory of team reasoning, an individual who reasons in the 'We' frame is aware of the 'I' frame too (as one that other players might use) but acknowledges only 'We' reasons. It seems that group identification involves something more than framing in the sense of variable frame theory: the group-identifier does not merely become aware of group concepts, she also becomes committed to the priority of group concepts over individual ones" (p.199).

In one of his unpublished papers (Bacharach, 1997), Bacharach allowed for the possibility of the existence of three frames: the I frame, the we frame and the 'S' (superordinate) frame. We and I are called simple frames: "players in them begin their reasoning with the two basic conceptualization of the situation, as 'what shall we do?' problem and 'what shall I do?' problem respectively" (p.5). An S frame is active when someone manages "during deliberation to see the problem from both the we and the I/she perspectives"(p.14). Although Bacharach allows for the existence of S, based on psychological attainments, he states that we and I perspectives cannot be held simultaneously: "Although we can switch self-identities rather easily, we appear to be unable to inhabit more than one at a time" (p.15). He assumes that I thoughts in the S frame generate a personal evaluation, whereas we thoughts generate a group evaluation. The solution concept in the model roughly states that the cooperative option is chosen by a player in S if it is the best in group evaluation and not worse than the other option in personal evaluation.

The S-frame intuition of the 1997 unpublished paper, however, disappeared in subsequent pieces of work. Later on, in developing VFT Bacharach faces the issue of integrability of frames. He says that normally frames are integrable:

> "It is easy to integrate frames which consist of classifiers such as shape, colour and position: we can easily see a mark as a triangle, as a blue triangle, as a blue triangle on the left,… on the other hand… a person can see the marks as letters and as geometric shapes, but not at the same time – you can't integrate these two perceptions" (2001 a, p.6).

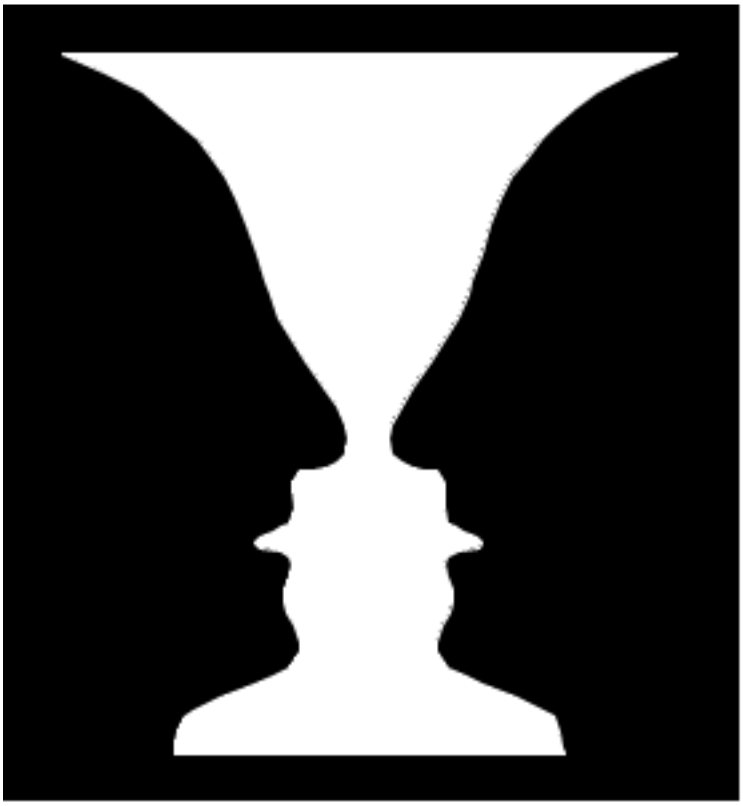


Figure 3.1: Rubin’s vase.

There exist frames, then, that are non-integrable. ‘I’ and ‘we’ frames appear non integrable in Bacharach’s words, and when this happens, “the agent may find herself vacillating between the judgments that she should do [two different actions]” (ib.). The idea that an agent can 'vacillate' between the two frames was so important for Bacharach that one of his (not realized) desires was to have Rubin's vase (Figure 3.1) on the front cover of the book[23] .

I shall suggest later that it is possible to take into account what Bacharach called ‘personal’ and ‘group’ evaluation, by reasoning in terms of deviation from an equilibrium and not in terms of frames. Or better, it is possible to do that, if we separate the two aspects of framing: how a frame might come to mind and how a person endorses a particular frame when she sees more than one frame.

[23] This comes from a personal communication with Robert Sugden, who inferred this desire by managing Bacharach's incomplete manuscript, which displayed Rubin's vase image on the first page.

3. The determination of mode of reasoning

To complete the theory, Bacharach needs to endogenize the determination of the mode of reasoning (this means the endogenization of ω). He tries to endogenize ω, because he sees the fact that ω is exogenous as a lacuna in his theory[24]. As we have seen in section 2.2, the theory of the determination of the mode of reasoning should not be a theory of rational choice. In his earlier works (Bacharach, 1997, 1999) he proposes that the possibility of team reasoning is related to having 'scope for cooperation' and to the 'harmony of interests'. Harmony is a non-strategic assesment of the game:

> "To endogenize ω, ..., one must show that the payoffs and other constitutive features of the basic game make collective identity salient or otherwise tend to induce team-thinking. The laboratory evidence is promising, as it suggests that group identification may be induced by the 'common problem' mechanism'. In addition, it is plausible that ω may be an increasing function of certain quantitative features of the payoff structure, such as 'scope for co-operation' and 'harmony of interest'" (1999, p.144).

A step forward on this topic has been made by Tan and Zizzo:[25] in their work there is an attempt to investigate the relationship between harmony of interests ('game harmony' for them), group identification and cooperation. They claim that game harmony is a good predictor of the extent of cooperation or conflict in games. They postulate that "game harmony increases cooperation by increasing the probability of team reasoning on the part of different players" (Zizzo 2004, p.20). Game harmony, defined as "a generic property describing how harmonious or disharmonious the interests of players are, as embodied in the payoffs" (Tan and Zizzo 2008, p. 3), is based on the correlation coefficient between payoff pairs - the Pearson

[24] "The unreliable team explanation of co-operative behavior I outlined in this paper contains an important lacuna. The distribution of agents over teams and the probability that they are active, are exogenous" (Bacharach 1999 p. 144).

[25] See Tan, J. and D. Zizzo (2008), Zizzo D. and Tan, J. (2007), and Zizzo D. (2004).

or Spearman correlation coefficient between the payoffs of the players for each state of the world for two player games[26]. This measure is the best existent proxy for what Bacharach has called 'the harmony of interest', and it is entirely derived from the payoffs of the game. It is a potential solution of Bacharach's problem of endogenization of ω. However, some of Bacharach's intuitions about vacillation cannot be expressed by the game harmony approach.

Bacharach also tries a second line in order to endogenize ω. This is the (strong) Interdependence Hypothesis, that roughly states: perceived interdependence prompts group identification. The perception of interdependence between two agents in a game is given by three factors:

- common interest (the agents have common interest in some s* over s, if both prefer s* to s, where s*, s are possible states of affairs, or, in a game, possible outcomes)
- copower (nobody can reach s* alone, but both can together)
- standard solution (basically the existence of a Nash equilibrium that realises s).

Basically, in Bacharach's interdepence hypothesis, if an outcome that can be reached by an individual way of reasoning (standard solution) is Pareto-dominated by another outcome that no individual can be sure of achieving on her own, but that can be achieved if all members of the group act in concert, there is space for group identification. The interdependence hypothesis uses I-reasoning as a default, makes use of opportunities for we-deviations that are good for 'us' and treats these opportunities as prompting we-reasoning. Interdependence fits with the intuition Bacharach had about vacillation between frames, but it seems to give an account only of we-deviation from I-thoughts. What about the opposite, that is from we-thoughts to I-thoughts? Bacharach offers only an informal conjecture about deviation from we to I: the 'double-crossing intuition'. Taking the most

[26] In general, in n-player games this measure is an average of Pearson (Spearman) correlation coefficients among payoff pairs.

famous game in terms of cooperation, the PD game, as an example, Bacharach says:

> “In a Prisoner’s Dilemma, players might see only, or most powerfully, the feature of common interest and reciprocal dependence which lie in the payoffs on the main diagonal” (Bacharach 2006, p.86).

If this happens, players do cooperate. But, it might be the case that

> “they might see the problem in other ways. For example, someone might be struck by the thought that her coplayer is in a position to double-cross her by playing D in the expectation that she will play C. This perceived feature might inhibit group identification” (ib).

Here Bacharach seems to have in mind some psychological process which inhibits group identity and which is not quite represented by his own concept of interdependence – the idea of ‘double-crossing’. The reason this idea does not fit his framework is that double-crossing is the incentive to act on individual reasoning when one believes the other is acting on team reasoning. And, what is more, double-crossing is a reason for a person who we-reasons to switch to the I-frame. A player, in order to recognize the ‘double-crossing’ threat, should be allowed to imagine herself in a we-frame, and then deliberating to cooperate, but at the same time she should use the I-frame by thinking that the other player would take advantage of her. In the first player’s conjecture, the other player too should use the we-frame in order to think that the first player could choose to cooperate, and, at the same time, she should use I-frame in order to think how to ‘double cross’ the first player. We may formalize what the statement 'i double-crosses j' means, i and j being the two players:

- [*Proposition P*]: i defects; i believes that j will cooperate; i believes that j believes that i will cooperate.

And 'j believes that i will double-cross j' means:

- j believes P.

It is now clearer that j's thoughts include: i acting on I-reasoning; i attributing we-reasoning to j; i attributing to j: attributing we-reasoning to i. In the theory of we-thinking the way in which a person reasons (I-mode or we-mode) is a consequence of the perceived frame. She may switch from the I-mode of reasoning to we-reasoning (if the we-frame comes to mind), or not. Bacharach, then, does not seem to take into account the possibility that once we are in the we-frame, we may switch to the I-mode of reasoning, or better, he allows the possibility of switching frame, but does not allow a person to be able to visualize switching frames. And this is why he cannot represent his 'double-crossing' intuition. It seems also, that when the we-frame is perceived, it is also perceived as the correct frame or dominant frame, so that once a person sees the world this way she cannot visualize going back to seeing it the other way (compare illusions, myths, lies – 'the scales fell from my eyes').

In order to complete Bacharach's theory in a more formal way, we need a model of vacillation, with deviation both from I to we and from we to I.

I shall present a first step in the next section, where I propose a representation of the double-crossing intuition.

4. Representing the 'double-crossing' intuition: reasoning in terms of deviations from equilibrium

In what follows, I shall present my analysis in terms of individual and collective rationality as two alternative ways of approaching a decision problem, and in particular I shall focus on reasons for deviating from an equilibrium. For simplicity I am considering two-player games, but the analysis could be easily extended to n-player games.

First of all, I suppose that the group utility function of a combination of actions is given by the mean of individual payoffs, as proposed by Bacharach (1999, 2006)[27] . A player who team reasons, first computes which is the best

[27] This formulation is the most used one in literature. I use it for concreteness, but any W which satisfies he Paretianness condition could be used without affecting the main conclusions of my analysis.

profile for the group[28] , and then, if this profile is unique[29] , he does his part in it. A player who ‘I’-reasons follows the standard theoretic predictions of game theory.

The basic idea is that a person may reason in the standard I-mode, or in we-mode, but she may have both frames (I and we) in mind (perhaps not at the same time, if the non-integrability hypothesis is correct, but vacillating between them). In standard game theory an equilibrium is defined as a combination of actions in which no player has anything to gain by changing unilaterally her own strategy. In we-reasoning theory, an equilibrium is instead defined as a combination of actions in which the whole group cannot gain anything by switching from this combination to another. Deviation is seen then as a test for the existence of an equilibrium, no matter if I or we-equilibrium.

To illustrate the principles underlying my proposal, I begin by considering some simple 2×2 symmetrical games. In each of these games there is at least one pure-strategy I-equilibrium and at least one pure-strategy we-equilibrium, and I consider only pure-strategy equilibria. I will then present my proposal in general in the next section. In the following tables, payoffs representing Nash equilibria are underlined and payoffs associated with we-equilibria are in italics.

Table 3.1: game A

	L	**R**
U	*3,3*	4,1
D	1,4	2,2

[28] In this version of the model, I am not taking account of the problems of 'unreliability' that Bacharach models by mean of circumspect team reasoning. The focus here is on vacillation, and at this stage I want to keep the model as simple as possible.

[29] In section 5 I will take in account also situations in which the optimal profile for the group is not unique.

Take, for example, the game A (table 3.1). The combination of actions (U, L) is a Nash equilibrium. Neither row player nor column player has reason to unilaterally deviate from that combination of actions. But the same combination is also a we-equilibrium: as a group both players cannot do better by switching to another combination[30].

Table 3.2: game B

	L	**R**
U	*2,2*	3,0
D	0,3	*2,2*

Game B shows a unique Nash equilibrium, (U, L) and two we-equilibria, (U, L) and (D, R), but only the (U, L) combination is an equilibrium at the same time for I and we-reasoners.

Table 3.3: game C

	L	**R**
U	*3,3*	1,1
D	1,1	2,2

Game C is strategically equivalent to a Hi-Lo game, and as is well known, it has two Nash equilibria, i.e. (U, L) and (D, R), but only one we-equilibrium, that is (U, L).

[30] The utility U for the group is 3 in the (U, L) combination, 2.5 in both (U, R) and (D, L), and 2 in (D, R).

Table 3.4: game D1

	L	**R**
U	*3,3*	1,4
D	4,1	2,2

Game D1 is a PD game, it has one Nash equilibrium (D, R) and one we-equilibrium (U, L), but these do not coincide.

Table 3.5: game D2

	L	**R**
U	*4,1*	0,0
D	*3,2*	1,3

Game D2, instead, has two We-equilibria (U, L) and (D, L) and two Nash equilibria (U, L) and (D, R), but only (U, L) is seen as an equilibrium from both I and we points of view. If an equilibrium survives both I and we deviation tests, it is particularly strong, in the sense that it allows for the existence of both ways of reasoning. At the same time such an equilibrium could be seen as a refinement when more than one - Nash or we - equilibrium exists. I shall call this equilibrium: I-we equilibrium. In game B, for example, there are two we-equilibria, but if we allow players to see the game endorsing both I and we concepts, this could help them to recognize that the (U, L) equilibrium is the prominent one, because it passes both deviation tests. In this case, having an I thought helps we-reasoners to select an equilibrium. But the opposite can happen as in the Hi-Lo game, where there are two Nash equilibria and we-thoughts can help I-reasoners to choose the (U, L) equilibrium.

This double test for deviation could also be seen in terms of deliberations, and not only as a method for testing the existence of an equilibrium. It can represent a model of transition between modes of reasoning, and as a component of a model of vacillation between them. The scheme in table 3.6 represents a possible way to classify the previous games in terms of deliberation, or vacillation.

Formally, if we define **I** as the set of I-solutions and **W** as the set of we-solutions, and we assume that they are non-empty, then the question "Are we always happy" means: is **I**⊆ **W**? And the question "Am I always happy" means: is **W**⊆ **I**?

Take for example game A: in this game, if I start to reason in the standard I-mode, we as a group will be happy with the result (U, L), i.e. we shall not want to deviate jointly from the I-reasoning 'solution'. Conversely, if I group identify, and then I look for the best solution for the group, I as an individual will be happy with the result, i.e. I will not want to deviate unilaterally from the we solution. So, in this game, the same result will be reached, independently of the particular way of reasoning. We may say that I or we-reasoning are observationally indifferent or equivalent, because they give the same result in terms of choice. Every game in which **I** and **W** coincide (**I**⊆ **W** and **W**⊆ **I**) belongs to this "Yes-Yes" category. But there could be different situations. Let us look at game B: in this case, if I start with the I-mode, there will be a unique Nash equilibrium (U, L), which is also one of the two possible (and indifferent) we-solutions. If I start with the I-mode, we shall then be happy with the result. If we group identify and we-reason, if we-reasoning gets us to (U, L), I am happy. But if it gets us to (D, R), I am unhappy. I may then turn to the I-mode of reasoning. In a vacillation process if, when reasoning in one mode, the conclusion is not endorsed by reasoning in the other mode, there is some tendency to switch to that other mode. So, in this case, the end of the vacillation process is the outcome (U, L), either by we-mode or by I-mode of reasoning. This result is observationally equivalent to I-reasoning but not to we-reasoning, because the latter allows (D, R). In this case the set I is a proper subset of **W** (**I**⊂**W**).

Start with WE: am I always happy with the results?

Start with I: are we always happy with the results?

<table>
<tr><th></th><th>YES</th><th>NO</th></tr>
<tr><th>YES</th>
<td><b>A</b>
<table>
<tr><th></th><th>L</th><th>R</th></tr>
<tr><td>U</td><td><b><i>3,3</i></b></td><td>4,1</td></tr>
<tr><td>D</td><td>1,4</td><td>2,2</td></tr>
</table></td>
<td><b>B</b>
<table>
<tr><th></th><th>L</th><th>R</th></tr>
<tr><td>U</td><td><b><i>2,2</i></b></td><td>3,0</td></tr>
<tr><td>D</td><td>0,3</td><td><b><i>2,2</i></b></td></tr>
</table></td></tr>
<tr><th>NO</th>
<td><b>C</b>
<table>
<tr><th></th><th>L</th><th>R</th></tr>
<tr><td>U</td><td><b><i>3,3</i></b></td><td>1,1</td></tr>
<tr><td>D</td><td>1,1</td><td><b>2,2</b></td></tr>
</table></td>
<td><b>D1</b>
<table>
<tr><th></th><th>L</th><th>R</th></tr>
<tr><td>U</td><td><b><i>3,3</i></b></td><td>1,4</td></tr>
<tr><td>D</td><td>4,1</td><td><b>2,2</b></td></tr>
</table>
<b>D2</b>
<table>
<tr><th></th><th>L</th><th>R</th></tr>
<tr><td>U</td><td><b><i>4,1</i></b></td><td>0,0</td></tr>
<tr><td>D</td><td><b><i>3,2</i></b></td><td><b>1,3</b></td></tr>
</table></td></tr>
</table>

x,x	We solution
x,x	I solution

Table 3.6: vacillation scheme

Game C, instead (the Hi-Lo game), is a mirror image of game B and will prompt we-reasoning: if we start by we-reasoning there will be a unique we-equilibrium (U, L), which is also one of the two possible Nash equilibria. So if we start with the we-mode, I will be happy with the result, and we shall not move from the (U, L) equilibrium. If instead I start with the I-mode we shall not always be happy: if the solution is (U, L), we shall be happy, but if it is (D, R) we shall not be happy, and we may turn to the we-mode of reasoning. Here we have **W⊂I**.

The game D1, the PD game, is the most interesting: if I start with I-reasoning, we shall not be happy (the Nash equilibrium is Pareto-dominated by the we-solution). But if we group identify the we-solution is not good for me (I would be better off by playing the other strategy). In this case there can be a continuous switching or vacillation from a frame to another: this could be an explanation of the empirical evidence on behaviour in PD games. In fact, in experiments on the PD game, we observe a rate of cooperation of about 50% (see Sally 1995). Following Bacharach's interdependence hypothesis, the PD, as we have seen, is one of the typical games that can lead to we-reasoning, although Bacharach himself was aware of the double-crossing threat. In the framework I have presented, the double-crossing intuition is taken into account, and this generates perpetual shifts between modes of reasoning, and then we-reasoning is only one of the two possible solutions. The PD game represents a category of games in which $\mathbf{I} \not\subseteq \mathbf{W}$ and $\mathbf{W} \not\subseteq \mathbf{I}$. In this particular case **I** and **W** are disjoint ($\mathbf{I} \cap \mathbf{W} = \emptyset$).

Game D2 belongs to the same category, but in this case the intersection between **I** and **W** is non empty ($\mathbf{I} \cap \mathbf{W} \neq \emptyset$): if I start with the I-mode and I-reasoning gets me to (U, L) we are happy; if I-reasoning instead gets us to (D, R) we are not happy. The same happens if we start with we-reasoning, because there are two we-equilibria and only one of them is also a Nash equilibrium.

Provided that **I** and **W** are non empty, the previous examples supply a complete classification of games in terms of intersection between I and we-equilibria.

This way of looking at a decision problem does not tell us which frame is more likely to appear. But, if a frame comes to mind, within this classification, we may see, depending on the kind of game the subject is facing, if the frame will be stable or not, or, in other words, we might see if that frame is an absorbing state in a model of transition or vacillation between frames.

In order to say something more about games with conflicting frames, in the next section I propose a formalization of the intuitions embedded in the previous classification of games.

5. A more formal vacillation model

The classification I proposed in the previous section represents a first step towards a generalization of Bacharach's model and intuitions. In the present section I show a possible way to generalize the previous results: I sketch a simple model based on I and we temptations to deviate from an equilibrium, and on a possible refinement of equilibria. In what follows I do not try to model explicitly how agents might reason about the mode of reasoning other agents are likely to adopt. I prefer at this stage to keep the model simple enough in order to provide the building blocks of a possible complete theory of we-reasoning. The model I shall propose can be thought as a model of how to endogenize ω (the probability the player is in the we-mode of reasoning). Once ω is determined, one can apply the standard 'circumspect team reasoning' procedure, as presented by Bacharach, in order to allow for unreliable teams. In the previous section I considered only pure-strategy equilibria but here I allow mixed strategies. A Nash equilibrium always exists in pure or mixed strategies and, as long as utility functions are continuous, it is possible to allow mixed-strategy equilibria. Being a maximum of a finite set, a we equilibrium always exists in pure strategies, but it is also possible to consider we-equilibria in mixed strategies. In this way we can be sure that I and W sets are non-empty.

We suppose that there are two players: 1, 2

S_1, S_2 are the strategy sets of players 1, 2, where $s_1 \in S_1$ and $s_2 \in S_2$ are the strategies chosen by players.

Let us define the following finite utility functions:

$U_1(s_1,s_2)$= 1's individual utility

$U_2(s_1,s_2)$= 2's individual utility

$W(s_1,s_2)$= we-utility.

Individual and group utilitiy functions have the characteristics specified in section 2.2., i.e. the individual utility is represented by a standard von Neumann - Morgenstern utility function, and group utlity is represented by a team utility function which satisfies the Paretianness condition.

Considering any candidate equilibrium $\left(s_1^*, s_2^*\right)$ from the viewpoint of player 1, we define:

- Own temptation to deviate $= \max_{s_1 \in S_1}\left[U_1\left(s_1, s_2^*\right) - U_1\left(s_1^*, s_2^*\right)\right] \equiv T_1\left(s_1^*, s_2^*\right)$
- Other's temptation to deviate $= \max_{s_2 \in S_2}\left[U_2\left(s_1^*, s_2\right) - U_2\left(s_1^*, s_2^*\right)\right] \equiv T_2\left(s_1^*, s_2^*\right)$
- Our temptation to deviate $= \max_{s_1 \in S_1, s_2 \in S_2}\left[W\left(s, s_2\right) - W\left(s_1^*, s_2^*\right)\right] \equiv T_W\left(s_1^*, s_2^*\right)$.

Temptations to deviate are necessarily greater than or equal to 0[31].

Now it is possible to define a Nash Equilibrium in terms of temptation to deviate. We have a Nash equilibrium when the following conditions hold:

$$\begin{cases} T_1\left(s_1^*, s_2^*\right) = 0 \\ T_2\left(s_1^*, s_2^*\right) = 0 \end{cases} \qquad [1]$$

A we-equilibrium is given when:

$$T_w\left(s_1^*, s_2^*\right) = 0 \qquad [2]$$

The next step is to define an I-we equilibrium: An I-we equilibrium is the intersection between **I** and **W**, and it exists when both conditions [1] and [2] hold.

For this reason, I-we equilibrium can be seen as:

(i) a refinement of Nash equilibirum

(ii) a refinement of we-equilibrium.

An I-we equilibrium helps to refine I-equilibria from a we point of view and we-equilibria from an I point of view, as we have seen in the classification in figure 2. Of course there can be more than one I-we equilibrium, just as there can be more than one Nash equilibrium.

[31] Take for example the game D1 (PD game): if we consider the profile (U, L) as a candidate equilibrium, we have that s_1 which maximizes $\left[U_1\left(s_1, L\right) - U_1\left(U.L\right)\right]$ is D; and then $T_1\left(U, L\right) = T_2\left(U, L\right) = \left[4 - 3\right] = 1$. If we consider (D,R) as a candidate equilibrium we have that s_1 which maximizes $\left[U_1\left(s_1, R\right) - U_1\left(D.R\right)\right]$ is D as well; and then $T_1\left(D, R\right) = T_2\left(D, R\right) = 0$.

But there could be cases, as in the PD game, in which an I-we equilibrium does not exist, because the conditions [1] and [2] can not both be met. When this happens, it is possible to generalize the analysis by developing Bacharach's intuition about vacillation. First of all, in order to do that, we need two candidate equilibria - a Nash and a we-equilibrium. So, if the game presents more than one Nash or we-equilibrium, we can imagine a refinement among them. For example, the following could be a refinement of Nash equilibrium:

(a) choose the Nash equilibrium which minimizes $h\left(T_W\left(s_1^*,s_2^*\right)\right)$, where $h\left(\ldots\right)$ is an increasing and finite (for finite *T_W*) function, with $h\left(0\right)=0$.

Or, in case of more than one we-equilibria, a refinement could be:

(b) choose the we-equilibrium which minimizes $l\left(T_1\left(s_1^*,s_2^*\right),T_2\left(s_1^*,s_2^*\right)\right)$ where $l\left(\ldots\right)$ is an increasing and finite (for finite T_1 and T_2) function, with $l\left(0,0\right)=0$.

If two or more equilibria are eligible in (a) and (b), we might postulate a random choice among them. This will not affect the remaining part of the analysis, as we shall see later[32].

We shall treat (a) and (b) as the candidate equilibria and we call:

- candidate Nash equilibrium $\equiv\left(s_1^*,s_2^*\right)$
- candidate we-equilibrium $\equiv\left(s_1^{**},s_2^{**}\right)$.

Now, we specify a solution in terms of the probability that the agent is acting in I-mode, i.e. the probability that $\left(s_1^*,s_2^*\right)$ is viewed as the solution by player 1 (or 2)[33], as well as the probability of its complementary event: the agent is acting in we-mode. We shall call these probabilities vacillation equilibrium. The probabilities used in vacillation equilibrium could be expressed in terms

[32] See footnote n. 34.

[33] The probability is the same for both players because we are considering temptations to deviate for both.

of Markov transition processes. In fact, Markov chains, with their properties, seem to represent the best candidate for a model of vacillation. Let us see how.

Suppose that the state space is $\Omega = \{I, W\}$,where I is I-reasoning and W is we-reasoning, and that $(X_0, X_1, \ldots)$ is the sequence of possible states of the process. If we call *p* the probability of transition from we to I, and *q* the probability of transition from I to we, we may define the transition matrix:

$$P = \begin{pmatrix} P(I,I) & P(I,W) \\ P(W,I) & P(W,W) \end{pmatrix} = \begin{pmatrix} 1-q & q \\ p & 1-p \end{pmatrix}$$

Let π_t be the probability distribution at *t*: $\pi_t = \left\{ \begin{matrix} \pi_{I,t} \\ \pi_{W,t} \end{matrix} \right\}$ is a vector whose components are the probability of I-reasoning at *t*, and the probability of we-reasoning at *t* for player *i*, where *i=1,2*. It is known that:

$$\pi_{t+1} = \pi_t P \qquad [3]$$

So that, for example, $\pi_{It+1} = \pi_{It}(1-q) + \pi_{Wt}p = \pi_{It}(1-q) + (1-\pi_{It})p$.

From equation [3], the probability distribution after n periods is given by $\pi_n = P^n \pi_0$. The eigenvalues of the matrix P are 1 and 1-(p+q), thus the matrix eigenvectors is given by:

$$X = \begin{bmatrix} p & 1 \\ q & -1 \end{bmatrix}$$

and the diagonal matrix is:

$$D = \begin{bmatrix} 1 & 0 \\ 0 & 1-(p+q) \end{bmatrix}.$$

We can derive *Pⁿ*:

$$P^n = XD^nX^{-1} = \frac{1}{p+q} \begin{bmatrix} p + q\{1+(p+q)\}^n & p - p\{1+(p+q)\}^n \\ q - q\{1+(p+q)\}^n & q + p\{1+(p+q)\}^n \end{bmatrix}$$

If $-1<1-(p+q)<1 \Leftrightarrow 0<p+q<2$, the model converges regardless of the initial probability distribution π_0 to

$$\lim_{n\to\infty}\pi_n = \frac{1}{p+q}\begin{bmatrix} p & p \\ q & q \end{bmatrix}\pi_0 = \begin{Bmatrix} \dfrac{p}{p+q} \\ \dfrac{q}{p+q} \end{Bmatrix} \equiv \begin{Bmatrix} \pi_I \\ \pi_W \end{Bmatrix}.$$

Suppose that *p*, the probability (for each player) of transition form we to I is an increasing function of $T_1\left(s_1^{**},s_2^{**}\right)$ and $T_2\left(s_1^{**},s_2^{**}\right)$ and that *q*, the probability of transition from I to we is an increasing function of $T_W\left(s_1^{*},s_2^{*}\right)$:

$$p = g\left(T_1\left(s_1^{**},s_2^{**}\right),T_2\left(s_1^{**},s_2^{**}\right)\right)$$

$$q = f\left(T_W\left(s_1^{*},s_2^{*}\right)\right)$$[34].

Then

$$\pi_I = pr\left[\text{player i is in I-mode}\right] = \frac{g\left(T_1\left(s_1^{**},s_2^{**}\right),T_2\left(s_1^{**},s_2^{**}\right)\right)}{f\left(T_W\left(s_1^{*},s_2^{*}\right)\right)+g\left(T_1\left(s_1^{**},s_2^{**}\right),T_2\left(s_1^{**},s_2^{**}\right)\right)}$$

and

$$\pi_W = pr[\text{player i is in we mode}] = \frac{f\left(T_W(s_1^*,s_2^*)\right)}{f\left(T_W(s_1^*,s_2^*)\right)+g\left(T_1(s_1^{**},s_2^{**}),T_2(s_1^{**},s_2^{**})\right)}$$

The probabilities π_I and π_W, where $\pi_W = 1-\pi_I$, are entirely derived from the temptations to deviate from equilibria, and represent a solution when an I-

[34] At this point it is clear that if we have more than one Nash or we-equilibrium which minimize respectively $f\left(T_W\left(s_1^{*},s_2^{*}\right)\right)$ or $g\left(T_1\left(s_1^{**},s_2^{**}\right),T_2\left(s_1^{**},s_2^{**}\right)\right)$, a random choice among them does not affect the resulting vacillation equilibrium. Because of the minimization criterion used in (a) and (b), the chosen Nash equilibrium must be equivalent in terms of $f\left(T_W\left(s_1^{*},s_2^{*}\right)\right)$, and the chosen we-equilibrium must be equivalent in terms of $g\left(T_1\left(s_1^{**},s_2^{**}\right),T_2\left(s_1^{**},s_2^{**}\right)\right)$. Due to the fact that the probabilities *p* and *q* are functions of $g\left(T_1\left(s_1^{**},s_2^{**}\right),T_2\left(s_1^{**},s_2^{**}\right)\right)$ and and $f\left(T_W\left(s_1^{*},s_2^{*}\right)\right)$, the values of p and q are the same, no matter which Nash or we-equilibrium we select.

we equilibrium does not exist. The model is then complete. In fact, π_W can be thought as ω in Bacharach's terms: the probability that a subject will group identify. In this way we have obtained a way to endogenize ω, remaining faithful to Bacharach intuitions. Starting from this point it is possible to apply Bacharach's analysis as illustrated of circumspect team reasoning. As I already said in the introduction, Bacharach offered two different proposals to complete his theory with the endogenization of I/we determination. One is game harmony, the other is the interdependence hypothesis. My approach is a way to formalize the latter. To see how these probabilities work in practice, take for example the following game:

Table 3.7: example

	a	b	c
a	10,10	0,0	0,15
b	0,0	9,9	0,10
c	15,0	10,0	1,1

In this game (a,a) is the unique we-equilibrium, and (c,c) is the unique Nash equilibrium. They do not coincide. As an illustration, suppose that:

$$f\left(T_W\left(s_1^*,s_2^*\right)\right)=T_W\left(s_1^*,s_2^*\right)$$

$$g\left(T_1\left(s_1^{**},s_2^{**}\right),T_2\left(s_1^{**},s_2^{**}\right)\right)=T_1\left(s_1^{**},s_2^{**}\right)+T_2\left(s_1^{**},s_2^{**}\right)$$

and that the group utility function W has the same form as the function used in section 4[35]. The vacillation model then gives:

[35] The group utility function of a combination of actions is given here by the mean of individual payoffs. A suitable property for *W*, but also for *f* and for *g* is to be invariant to positive affine transformations of the utlitities. Following Bacharach this could be made possible by embedding in *W* the principles of Nash's axiomatic bargaining theory (trying to find the equivalent of the disagreement reference point). Another possibility is to use utilities which are ratio scale measurable, but, as it has been demonstrated by Tsui (1997), they allow weak Pareto dominance principle only.

$$\pi_I = pr\left[1 \text{ plays c}\right] = \frac{5+5}{5+5+9} = \left(\frac{10}{19}\right)$$

$$\pi_W = pr\left[1 \text{ plays a}\right] = \frac{9}{5+5+9} = \left(\frac{9}{19}\right)$$

This game belongs to category D1 (the category which includes the PD game) following the scheme proposed in the previous section: this means that the deliberation process leads to a continuous switching between frames. The structure of payoffs, however, and then the relative strenght of I and we temptations to deviate, allow us to infer that the strategy associated with the I solution is more likely to be selected by each player.

Obviously my model does not represent a complete theory of how we-thinking arises: the psychological bases of the process, such as group identification, salience deriving from how strategies are labelled, etc., are also important, but at the same time they are difficult to model in a game theoretical form. In this paper, following Bacharach I have concentrated on a mechanism, which can be entirely derived from payoff functions.

A further step of research could be a comparison, in terms of predictions between the game harmony measure, the vacillation model, and other behavioural predictions about cooperation in games. It is worth noticing that some games, for example Game A (table 3.1 section 4) and Game D1 (table 3.4 section 4) have the same game harmony measure (in this case -0.8), but they are different in terms of reasoning about deviations, and therefore in terms of the vacillation model. In Game A, I and we-modes of reasoning are observationally equivalent because both lead to the same solution. In Game D1, instead, there will be a continuous switching between the I and we modes of reasoning, given that $\pi_I = \pi_W = 0.5$. At the same time, by slighlty changing the payoffs of Game A, harmony will change but not the way of reasoning. Let us see an example.

Table 3.8: second example

	L	R
U	*4,4*	3,1
D	1,3	2,2

The game in table 3.8 belongs to category A, but now the game harmony has become positive: it is 0.2. One of the reasons for this differences is that my approach, unlike that of game harmony, is based on strategic reasoning. These are only examples, but they show that there is space for a comparison in terms of behavioral predictions between my proposal and the game harmony measure, as well as other behavioral predictions, deriving from theories of social preferences, which do not deal with we-reasoning.

6. Conclusions

In this paper I have analysed and extended Bacharach's theory of we-thinking. This is a very well developed formal theory of games with I-reasoning and we-reasoning, with the mode of reasoning taken as given. A fundamental feature of the theory is that the mode of reasoning is prior to rational choice. So, as Bacharach himself recognises, to complete the theory there has to be a model of which mode of reasoning is used by the agents. This part is less formal and less developed by Bacharach, although he offers many intuitions and suggestions. In particular I have described two approaches Bacharach attempted to use: the harmony approach, developed by Zizzo and Tan, and the interdependence idea, which contains an underdeveloped intuition about vacillation between frames, and is only one way - I to we- in its formal presentation, but it seems naturally two way - I to we and we to I - as we see in the double-crossing intuition.

I have proposed a way in which the double-crossing intuition may be taken into account: reasoning about deviation from equilibrium, where equilibrium is seen both from an I and from a we point of view. I have presented a classification of games, based on reasons to deviate from an equilibrium (I or we), suggesting that an I-we equilibrium could be represented by the intersection between I and we equilibria. However, in some games the intersection can be empty, and this leaves space for vacillation. In order to determine which mode of reasoning is more likely to be chosen in these cases, I have presented a more formal model, based on temptation to deviate from an equilibrium and a vacillation equilibrium which can be induced by Markov transition processes. I have argued that this approach develops Bacharach's theory in a way that is faithful to his intuitions and that makes we-reasoning more easily usable in game theory.

Chapter 4
The roles of level-k and team reasoning in solving coordination games

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'[36].

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. 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

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

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[37]. 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

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

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.

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[38]. 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.

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.

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

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)[39].

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 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

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

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 the slices, which in our experiment are represented as three circles, 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 Table 4.1. The parameters x, y, v, w are always strictly positive and satisfy y≥x and {v, w} ≠ {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.

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

Table 4.1: Pie game's payoff matrix

Figure 4.1 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.

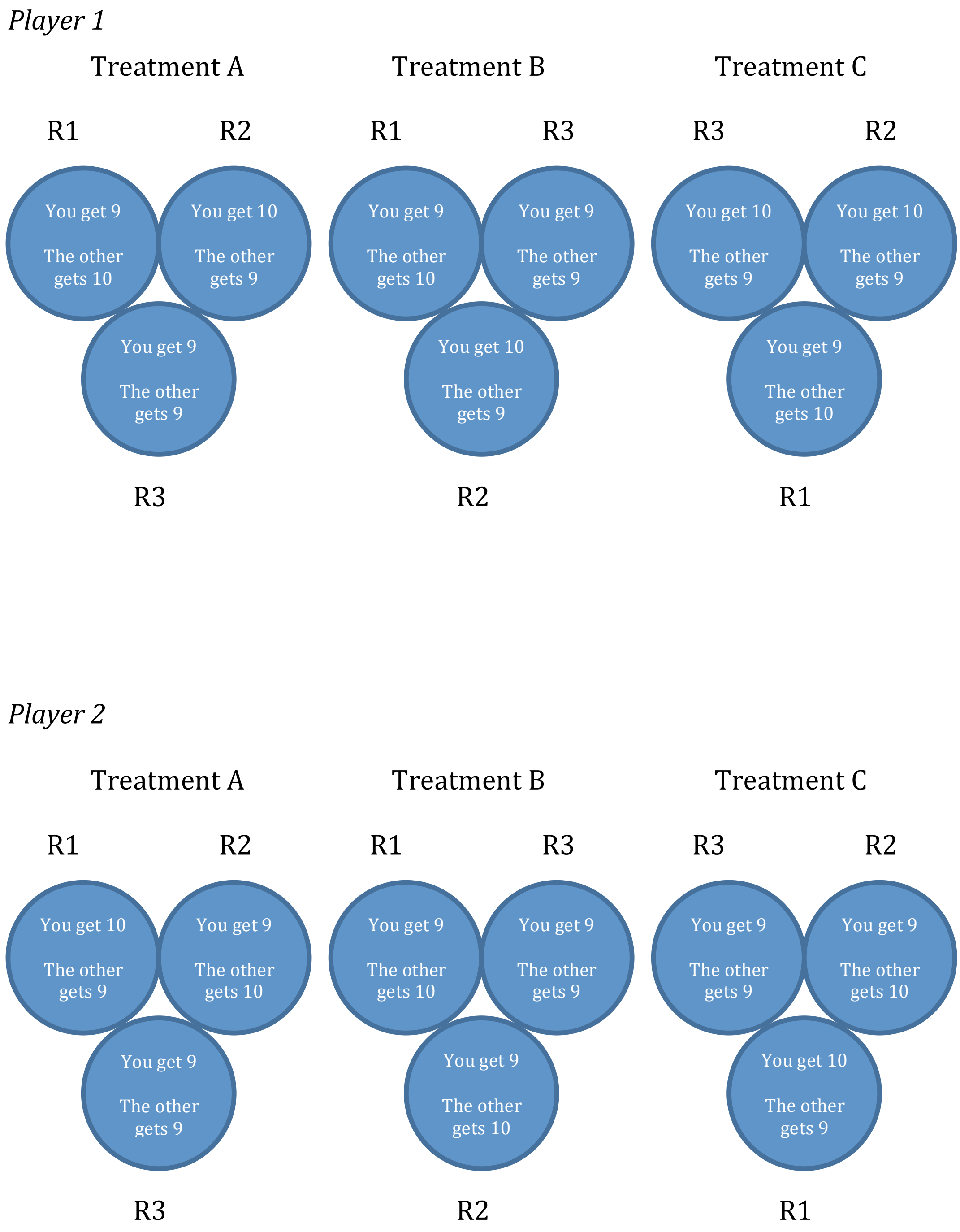

Player 1
Treatment A
R1
R2
You get 9 The other gets 10
You get 10 The other gets 9
You get 9 The other gets 9
R3
Treatment B
R1
R3
You get 9 The other gets 10
You get 9 The other gets 9
You get 10 The other gets 9
R2
Treatment C
R3
R2
You get 10 The other gets 9
You get 10 The other gets 9
You get 9 The other gets 10
R1
Player 2
Treatment A
R1
R2
You get 10 The other gets 9
You get 9 The other gets 10
You get 9 The other gets 9
R3
Treatment B
R1
R3
You get 9 The other gets 10
You get 9 The other gets 9
You get 9 The other gets 10
R2
Treatment C
R3
R2
You get 9 The other gets 9
You get 9 The other gets 10
You get 10 The other gets 9
R1

Figure 4.1: The three treatments (Game G1).

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 4.2.

Table 4.2. Outcomes properties.

Game	Payoff			R1/R2	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

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. $v \geq x,y$ and $w \geq x,y$ 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[40].

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.

[40] Note that because of the condition $\{v, w\} \neq \{x, y\}$, 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.

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

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)[41].

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 x≤y. Then a sufficient condition for R3 to be team optimal is U(w,w)>U(y,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 4.1) the lowest value of w/y is 8/10. Thus R3 is (very strongly) team optimal in every game.

3.2.2 Level-k predictions

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

[41] See Gold (2012) for a review of Utility group function properties.

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 4.3).

Table 4.3: 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	R1	R2
Player 2 choice	R1, R2, or R3	R1	R2	R1

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 4.4 for game 1.

Table 4.4: Level-K theory's prediction in Game 1 when L0 follows payoff salience.

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

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 ≥ 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 ≥ 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 ≥ 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 4.5. 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 4.5. 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

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 ≠ w and x ≠ 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 4.6.

Table 4.6. 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

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[42]. 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.

[42] Because of a mistake in programming the software, we run the first three sessions with game 1 and 11 slighlty different from what we had planned. This results in having 194 players for games 2-10, and only 140 for games 1 and 11. We decided to keep the data on games 2-10 we colleted in the first three sessions, because the design of the experiment was the same.

4 Results

Table 4.7 reports the distribution of R1, R2 and R3 choices across treatments in the 11 games.

Table 4.7. 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***
* 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%

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 3).

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 3)[43].

We also test the null hypothesis that subjects chose at random. The null hypothesis is rejected in all the games except G10[44].

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 4.7).

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.[45]

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

[43] 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.

[44] 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

[45] 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.

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.[46]

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 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

[46] 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.

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 4.2) reported by Crawford et al. (2008), which is very similar to game G3.

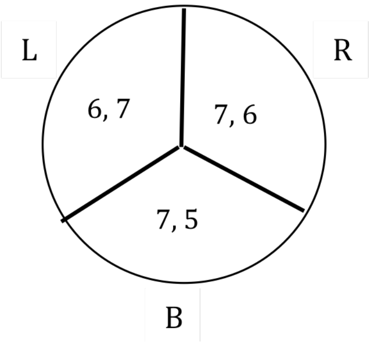


Figure 4.2. Crawford's et al. (2008) AM4 Game.

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.

Appendix 2. 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 3. Distribution of choices across treatments

GAME G1		Treat.A	Treat.B	Treat.C
	R1/R2	3	7	8
P1				
	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			
	R1/R2	5	5	8
P2				
	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
	R1/R2	1	0	0
P1				
	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			
	R1/R2	0	0	0
P2				
	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
	R1/R2	29	37	26
P1				
	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			
	R1/R2	30	38	27
P2				
	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
	R1/R2	6	10	4
P1	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			
	R1/R2	8	5	6
P2				
	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. Rat	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. Rat	37	43.5	44.4

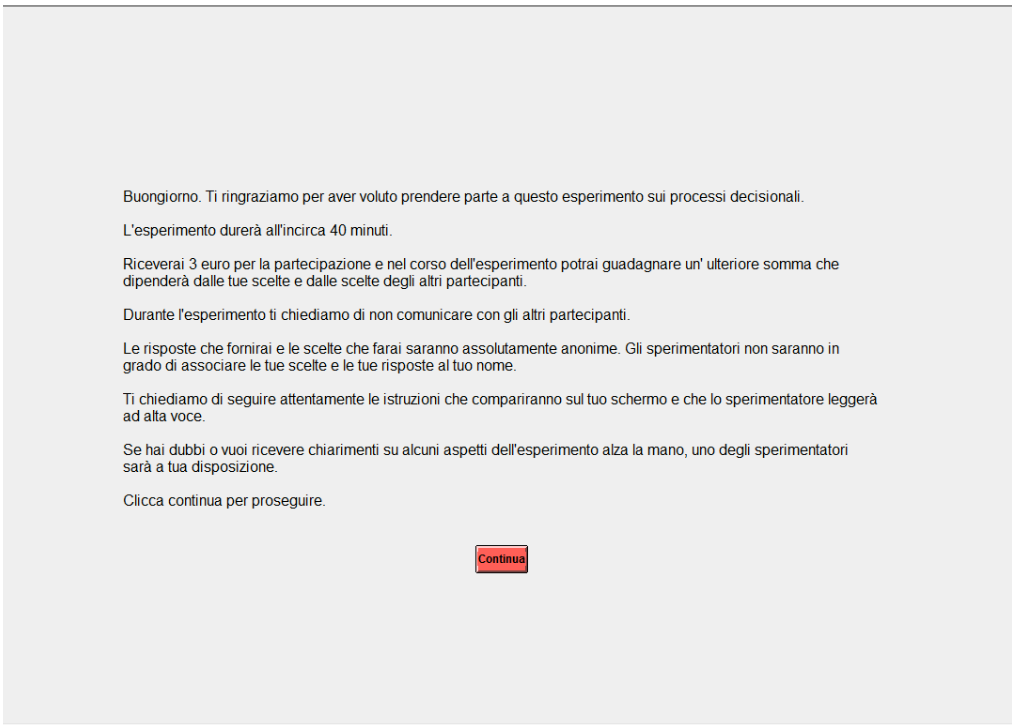
	GAME G9			
		Treat.A	**Treat.B**	**Treat.C**
P1	R1/R2	8	5	6
	R3	25	34	19
	TOT	**33**	**39**	**25**
	Binomial two sided test (n=R1+R2+R3 choices; k=R3 choices; p=0.33)	p<0.01	p<0.01	p<0.01
	H0 : same distr. across treatments; Pearson chi2(2) = 1.89 p>0.05			
P2	R1/R2	10	10	14
	R3	23	29	11
	TOT	**33**	**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**
P1	R1/R2	21	28	20
	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			
P2	R1/R2	21	27	18
	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**
P1	R1/R2	2	6	2
	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			
P2	R1/R2	3	2	4
	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

Appendix 4 Instructions

1. Instructions



Buongiorno. Ti ringraziamo per aver voluto prendere parte a questo esperimento sui processi decisionali.

L'esperimento durerà all'incirca 40 minuti.

Riceverai 3 euro per la partecipazione e nel corso dell'esperimento potrai guadagnare un' ulteriore somma che dipenderà dalle tue scelte e dalle scelte degli altri partecipanti.

Durante l'esperimento ti chiediamo di non comunicare con gli altri partecipanti.

Le risposte che fornirai e le scelte che farai saranno assolutamente anonime. Gli sperimentatori non saranno in grado di associare le tue scelte e le tue risposte al tuo nome.

Ti chiediamo di seguire attentamente le istruzioni che compariranno sul tuo schermo e che lo sperimentatore leggerà ad alta voce.

Se hai dubbi o vuoi ricevere chiarimenti su alcuni aspetti dell'esperimento alza la mano, uno degli sperimentatori sarà a tua disposizione.

Clicca continua per proseguire.

Welcome to this experiment on decisional processes and thank you for participating in it.

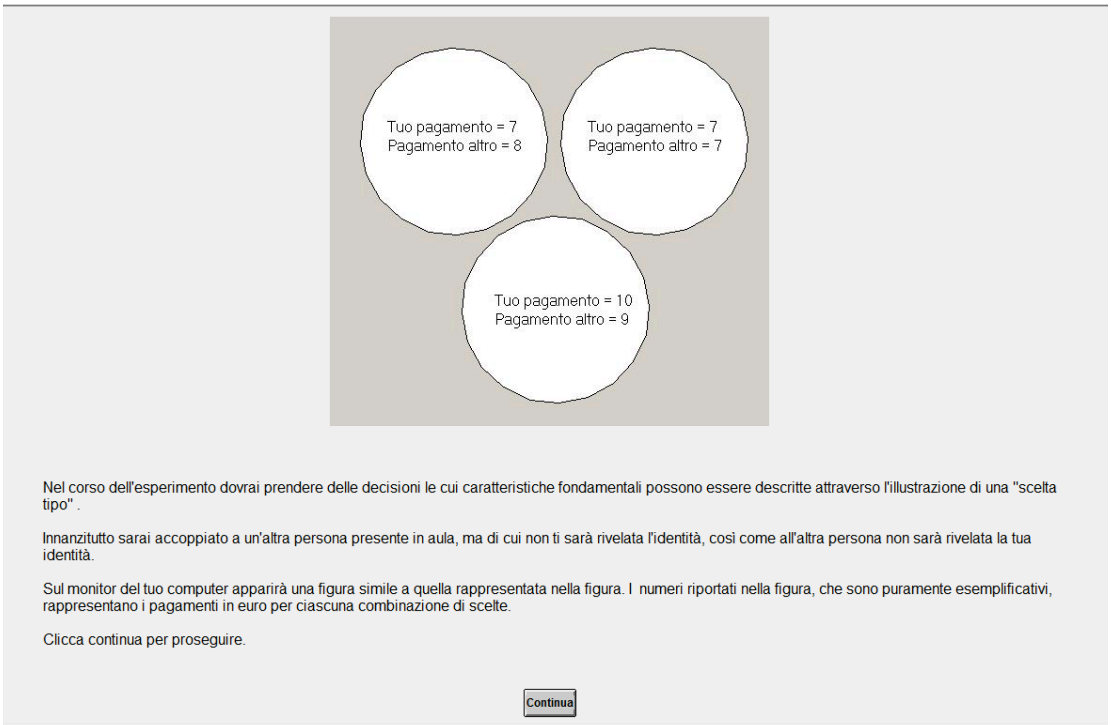
The experiment will last approximately 40 minutes.

You will receive 3 euro for your participation. You can gain more money, depending on your choices and on choices of other participants.

Your answers and your choices will be completely anonymous. The experimenters are not able to associate your choices and your answers to your name.

We ask you to pay attention to the instructions that will appear on your screen. They will be read aloud by one of the experimenters.

If you have any doubts or questions about anything related to the experiment raise your hand: one of the experimenters will come.



Nel corso dell'esperimento dovrai prendere delle decisioni le cui caratteristiche fondamentali possono essere descritte attraverso l'illustrazione di una "scelta tipo" .

Innanzitutto sarai accoppiato a un'altra persona presente in aula, ma di cui non ti sarà rivelata l'identità, così come all'altra persona non sarà rivelata la tua identità.

Sul monitor del tuo computer apparirà una figura simile a quella rappresentata nella figura. I numeri riportati nella figura, che sono puramente esemplificativi, rappresentano i pagamenti in euro per ciascuna combinazione di scelte.

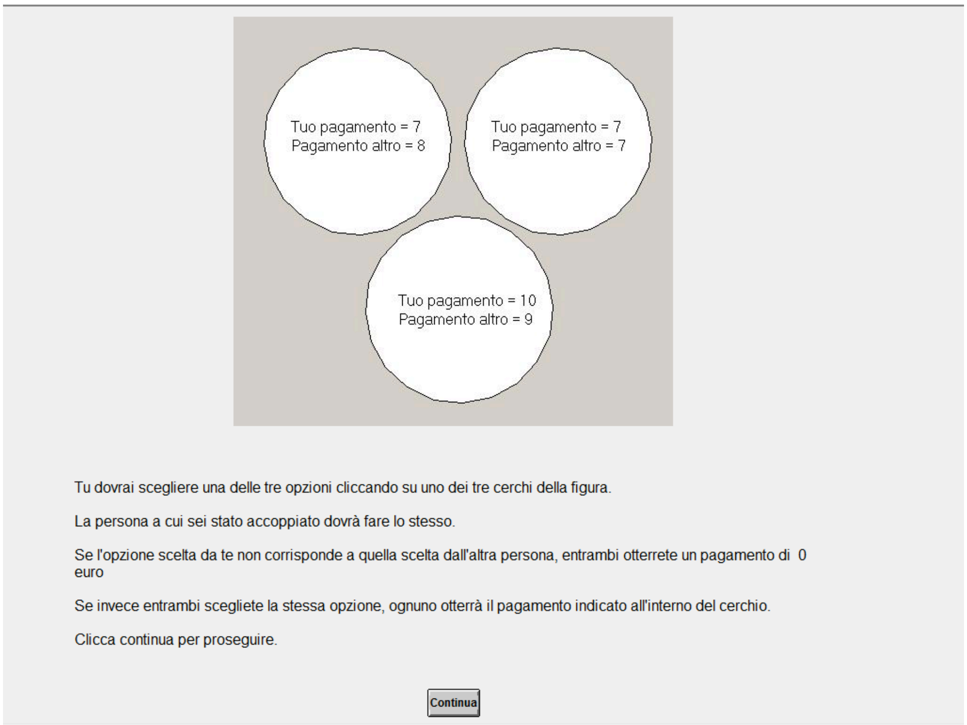
Clicca continua per proseguire.

During the experiment you will be asked to make choices. The main characteristics of the choices can be described through an example.

First of all you will be matched with another person in this room. You will never know the identity or the other person, nor will he/she know your identity.

On the screen of your computer will appear a figure similar to the one you are now seeing above. The numbers on the figure, which are only examples, represent the payments in euro for each combination of choices.

Click on 'continue' to proceed.



Tu dovrai scegliere una delle tre opzioni cliccando su uno dei tre cerchi della figura.

La persona a cui sei stato accoppiato dovrà fare lo stesso.

Se l'opzione scelta da te non corrisponde a quella scelta dall'altra persona, entrambi otterrete un pagamento di 0 euro

Se invece entrambi sceglete la stessa opzione, ognuno otterrà il pagamento indicato all'interno del cerchio.

Clicca continua per proseguire.

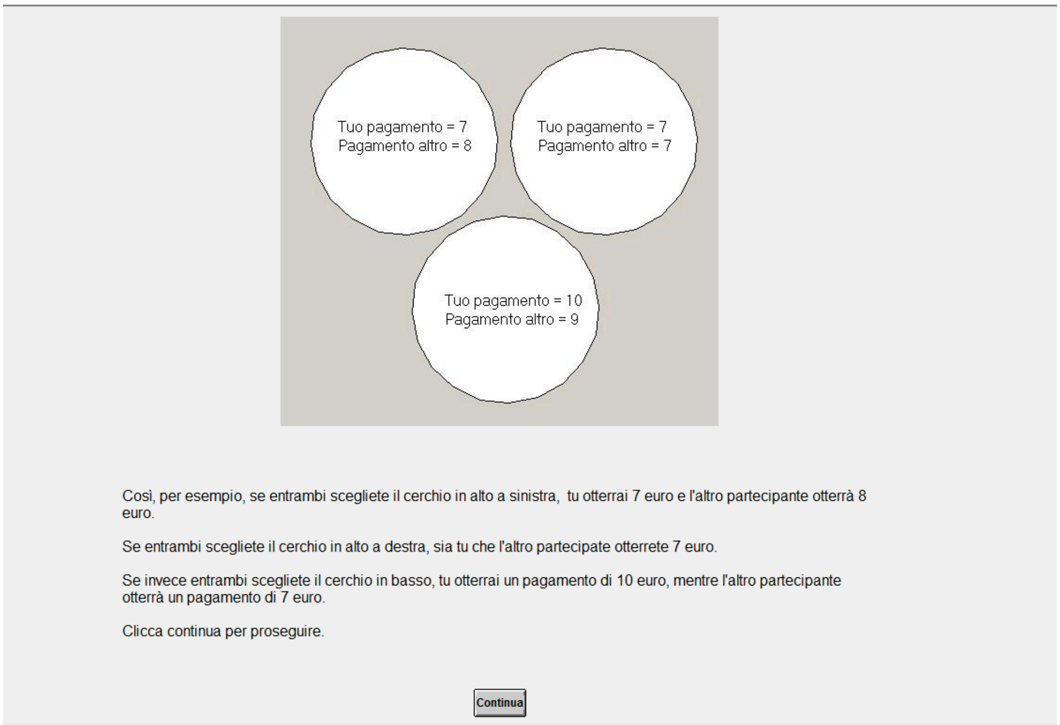
You must choose one of the three options by clicking on one of the three circles on the figure.

The person you are matched with must do the same.

If your choice does not correspond to the choice of the other person, you will both get 0 euro.

If both of you choose the same option, each of you will obtain the payment which is in the circle.

Click on ‘continue’ to proceed.



Così, per esempio, se entrambi scegliete il cerchio in alto a sinistra, tu otterrai 7 euro e l'altro partecipante otterrà 8 euro.

Se entrambi scegliete il cerchio in alto a destra, sia tu che l'altro partecipate otterrete 7 euro.

Se invece entrambi scegliete il cerchio in basso, tu otterrai un pagamento di 10 euro, mentre l'altro partecipante otterrà un pagamento di 7 euro.

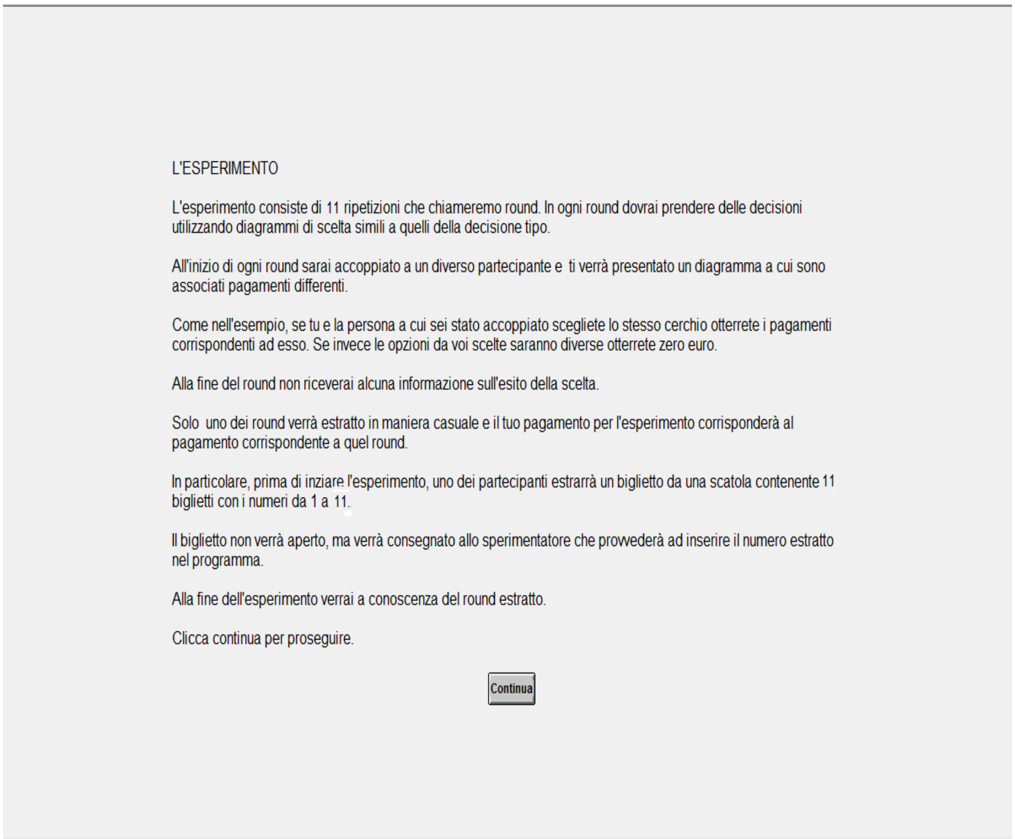
Clicca continua per proseguire.

So, if both of you choose the top left circle, you will get 7 euro and the other will obtain 8 euro.

If both of you choose the top right circle, you will get 7 euro, the other will get 7 euro.

If both of you choose the bottom circle, you will obtain 10 euro, and the other will get 7 euro.

Click on 'continue' to proceed.



L'ESPERIMENTO

L'esperimento consiste di 11 ripetizioni che chiameremo round. In ogni round dovrai prendere delle decisioni utilizzando diagrammi di scelta simili a quelli della decisione tipo.

All'inizio di ogni round sarai accoppiato a un diverso partecipante e ti verrà presentato un diagramma a cui sono associati pagamenti differenti.

Come nell'esempio, se tu e la persona a cui sei stato accoppiato scegliete lo stesso cerchio otterrete i pagamenti corrispondenti ad esso. Se invece le opzioni da voi scelte saranno diverse otterrete zero euro.

Alla fine del round non riceverai alcuna informazione sull'esito della scelta.

Solo uno dei round verrà estratto in maniera casuale e il tuo pagamento per l'esperimento corrisponderà al pagamento corrispondente a quel round.

In particolare, prima di inziare l'esperimento, uno dei partecipanti estrarrà un biglietto da una scatola contenente 11 biglietti con i numeri da 1 a 11.

Il biglietto non verrà aperto, ma verrà consegnato allo sperimentatore che provvederà ad inserire il numero estratto nel programma.

Alla fine dell'esperimento verrai a conoscenza del round estratto.

Clicca continua per proseguire.

The experiment.

The experiment consists of eleven rounds. In each round you must make decisions on figures similar to the ones you have seen before.

At the beginning of each round you will be matched with a different person, and you will see a figure with different payments.

As in the previous example, if you and the other person choose the same circle, you will get the corresponding amount. If you and the other person choose a different circle the payment will be 0 euro.

At the end of each round you will not receive any feedback on the results.

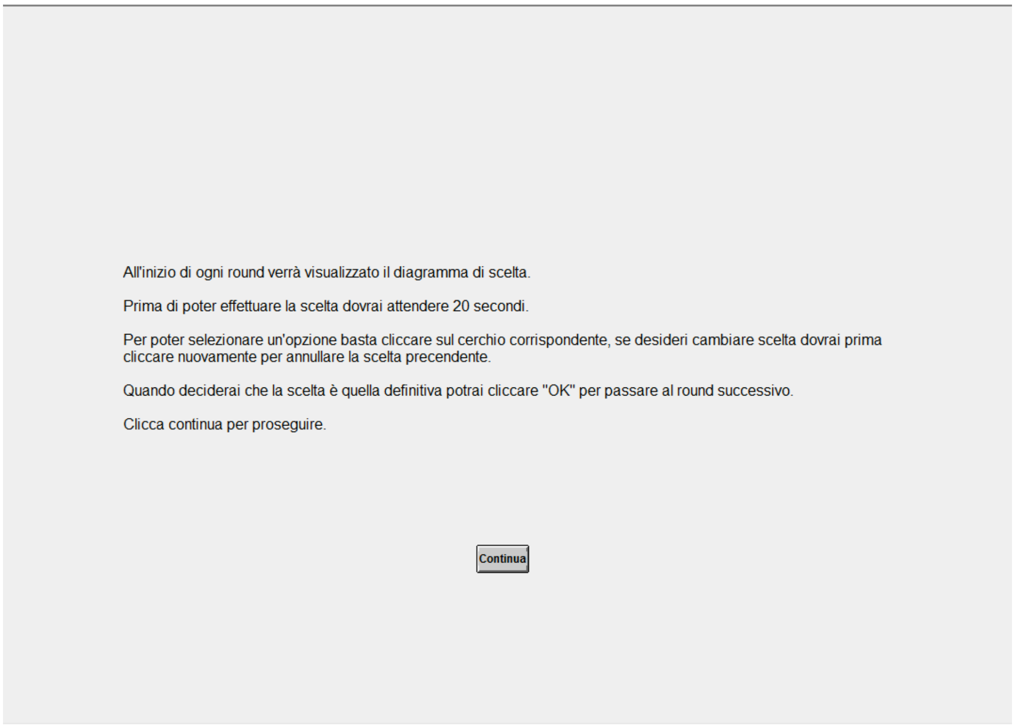
At the end of the experiment, one of the rounds will be randomly selected, and you will receive the corresponding payment associated to that round.

In particular before starting the experiment one participant will draw a ticket from a box containing 10 tickets numbered from 1 to 11.

The ticket (without being opened) will be given to the experimenter.

At the end of the experiment you will know the selected round.

Click on ‘continue’ to proceed.



All'inizio di ogni round verrà visualizzato il diagramma di scelta.

Prima di poter effettuare la scelta dovrai attendere 20 secondi.

Per poter selezionare un'opzione basta cliccare sul cerchio corrispondente, se desideri cambiare scelta dovrai prima cliccare nuovamente per annullare la scelta precendente.

Quando deciderai che la scelta è quella definitiva potrai cliccare "OK" per passare al round successivo.

Clicca continua per proseguire.

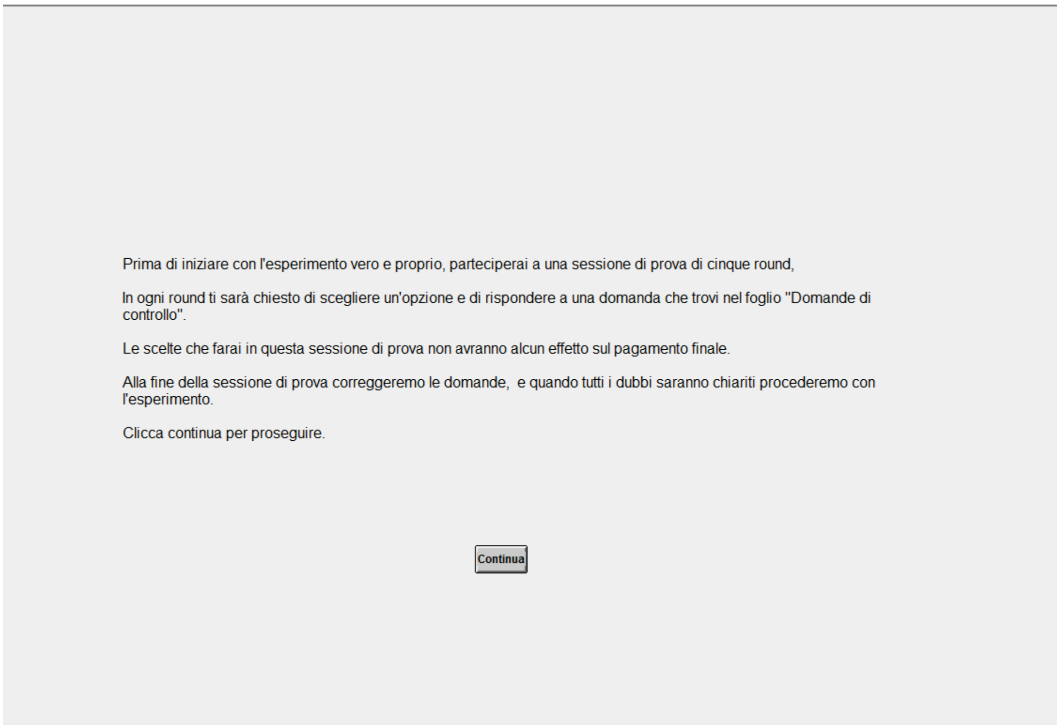
At the beginning of each round the figure with three circles will be shown.

Before making your choice you must wait for 20 seconds.

In order to select a choice it is sufficient to click on the selected circle. If you want to change your choice, you must click again on the selected choice in order to deselect it, then you can click on the other choice.

When your decision is definitive you must click on the ‘ok’ button, in order to go to the following round.

Click on ‘continue’ to proceed.



Prima di iniziare con l'esperimento vero e proprio, parteciperai a una sessione di prova di cinque round,

In ogni round ti sarà chiesto di scegliere un'opzione e di rispondere a una domanda che trovi nel foglio "Domande di controllo".

Le scelte che farai in questa sessione di prova non avranno alcun effetto sul pagamento finale.

Alla fine della sessione di prova correggeremo le domande, e quando tutti i dubbi saranno chiariti procederemo con l'esperimento.

Clicca continua per proseguire.

Before starting the experiment, you will participate in a trial session with 5 rounds.

In each round you will be asked to choose an option and to answer a question, which is on the sheet on your table.

The choices made in this session will not affect your final payment.

At the end of this session we shall correct the questions and we shall clarify any doubts. After that we shall proceed with the experiment.

Click on ‘continue’ to proceed.

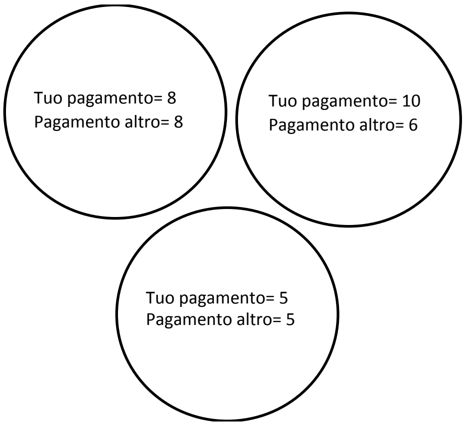
2. Control questions

We used 5 control question, one for each trial game. They were on a separated sheet, and we asked subjects to write their answers there.

This is an example:

CONTROL QUESTIONS **PC NO: ………..**

ROUND 1



If you choose the top-left circle and the other participant chooses the top-right circle,

Your payment is =………..

The other participant's payment is =……..

The other cases were:

You: bottom

The other: top-left

You: top-left

The other: top-left

You: top-left

The other: bottom

You: bottom

The other: bottom

References

Anderson, E. (2001): “Unstrapping the straitjacket of ‘preference’: a comment on Amartya Sen’s contribution to Philosophy and Economics”, Economics and Philosophy, 17: 21-38.

Antoci A., Sacco P. and Zarri L. (2004): “Coexistence of Strategies and Culturally- Specific Common Knowledge: An Evolutionary Analysis”, Journal of Bioeconomics, vol. 6:165-194.

Axerlod R. (1984): *The evolution of Cooperation*, Basic Books.

Bacharach, M. (1993): "Variable Universe Games," in *Frontiers of Game Theory*, ed. by K. G. Binmore, A. P. Kirman, and P. Tani. Massachusetts, MIT Press.

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

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

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

Bacharach, M. (2001): "Framing and Cognition in Economics: The Bad News and the Good," ISER Workshop, Cognitive Processes in Economics, available at http://cess-wb.nuff.ox.ac.uk/documents/mb/lecnotes.pdf.

Bacharach, M. (2006): *Beyond Individual Choice*, Princeton University Press, Edited by N. Gold, and R. Sugden.

Bacharach, M., Bernasconi, M. (1997): "The Variable Frame Theory of Focal Points: An Experimental Study", in Games and Economic Behavior, 19 (1): 1-45.

Barsdley, N., Mehta J., Starmer C. and Sugden, R. (2010): "Explaining Focal Points: Cognitive Hierarchy Theory versus Team Reasoning", The Economic Journal 120 (March): 40-79.

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

Becchetti, L., Degli Antoni G., and Faillo, M. (2009): "Common reason to believe and framing effect in the team reasoning theory: an experimental approach", Econometica Working Paper series, n.15, November.

Benn, S. (1978): "The problematic rationality of political participation" in Benn, *Political Participation*, Canberra: Australian National University Press: 61-68.

Bergstrom T. C. and Stark O. (1993): "How Altruism Can Prevail in an Evolutionary Environment", The American Economc Review, vol. 83 (2): 149-155.

Bicchieri C. (1997): "Learning to cooperate", In Bicchieri, Jeffrey, Skyrms (ed.) (1997): *The dynamics of norms*, Cambridge, Cambridge University Press.

Bicchieri C. (2006): *The Grammar of Society: The Nature and Dynamics of Social Norms*, Cambridge, Cambridge University Press.

Balkenborg, D. (1993): "Strictness, Evolutionary Stability and Repeated Games with Common Interests", CARESS, Working Paper: 93-20.

Binmore K. (2005): *Natural Justice*, Oxford University Press, New York.

Binmore K. (2006): "Why do people cooperate?", Politics, Philosophy and Economics, vol. 5 (1): 81-96.

Bolton G. and Ockenels A. (2000): “ERC: A theory of equity, reciprocity and competition”, American Economic Review, Mar 2000, vol. 90 (1): 166-194.

Bomze I. (1983): “Lotka-Volterra Equation and Replicator Dynamics: A Two-Dimensional Classification”, Biological Cybernetics, vol. 48: 201- 211.

Boyd R. and Richerson P. (1985): *Culture and the Evolutionary Process*, Chicago, The University of Chicago Press.

Bowles S., Gintis H. (2004): "The evolution of strong reciprocity: cooperation in a heterogeneous population." Theoretical Population Biology, vol. 65:17- 28.

Bruni, L. (2008): *Reciprocity, altruism and the civil society*, Routledge.

Bruni L., Smerilli A. (2010): “Cooperation and diversity”, Munich Working Paper Series, n. 20564.

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): 861–98.

Casajus A. (2000): “Focal Points in Framed Strategic Forms”, Games and Economic Behavior, 32: 263-291.

Collard D. (1978), *Altruism and Economics*, Oxford, Martin Robertson.

Collard D. (1983): “Economics and philanthropy: a comment”, Economic Journal, vol. 93: 637-8.

Colman, A.M., Pulford B.D. 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, 98 (4): 1443-1458.

Darwin, Charles (1859): *On the Origin of Species by Means of Natural Selection*, John Murray, Albemarle, Street, London.

Davis J. (2001): The intersubjectivity in economics, ed. by, Routledge, London.

Davis, J. B. (2009): “Two relational conceptions of individuals: teams and neuroeconomics” in A Research Annual, Research in the History of Economic Thought and Metodology series, vol. 27 – A: 1-21.

Dufwenberg M., Kirchsteiger G. (2004): “A theory of sequential reciprocity”, Games and economic behaviour, vol. 47: 268-298.

Elster J. (1989): *The cement of society*, Cambridge University Press.

Falk A., Fischbacher U, (2005): “Modeling Strong Reciprocity” in Gintis, H., Bowles, S., Boyd, R., Fehr, E. (2005), ed. by, *Moral sentiments and material interest. The foundations of Cooperation in Econnomic Life*, MIT Press.

Fehr E., Fischbacher, U. (2005): “The economics of strong reciprocity”, in Gintis e al. (2005).

Fehr E., Gachter S. (2000): “Fairness and Retaliation: The Economics of Reciprocity”, Journal of Economic Perspectives, 14:159-181.

Fehr E., Schmidt K. (1999): “A theory of fairness, competition, and cooperation”, *Quarterly Journal of Economics* vol. 114: 817-868.

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

Gauthier, D. (1975): “Coordination” in Dialogue, 14:195-221.

Gauthier, D (1986): *Moral by agreements*, Oxford, Oxford University Press.

Gilbert, M. (1989): *On Social Facts*. Routledge.

Gintis H. (2003): "A critique of team and Stackelberg reasoning", Behavioral and Brain Sciences, vol. 26: 161-162

Gintis H (2004): "Modeling Cooperation Among Self-Interested Agents: A Critique", The Journal of Socio-Economics, 33:311-322.

Gintis H (2009): *Game theory evolving*, New Jersey, Princeton University Press.

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, N., and Sugden R. (2008): "Theories of Team Agency," in *Rationality and Commitment*, ed. by F. Peter, and S. Schmidt: Oxford University Press.

Guala, F., Mittone, L. and Ploner, M. (2009): "Group membership. team preferences and expectations", CEEL working papers.

Hargreaves Heap S. (1997): "When norms influence behaviour: Expressive Reason and its consequences", mimeo, University of East Anglia.

Hargreaves Heap S. and Varoufakis Y. (2004): *Game theory. A critical text*, London, Routledge.

Harsanyi, J. C. and Selten, R. (1998): *A general theory of equilibrium selection in games*. Cambridge, MA: MIT Press.

Heckathorn D. (1996): "The dynamics and dilemmas of collective action", American Sociological Review, vol. 61:250-277.

Heller W. B., Sieberg K. K. (2010): "Honor among thieves: Cooperations as a strategic response to functional unpleasantness", European Journal of Political Economy, vol. 26:351-382.

Hirshleifer J., Martinez Coll J. (1991): "The limits of reciprocity", Rationality and Society, vol. 3:35-64

Ho, T., Camerer C., WeigelT K., (1998): "Iterated Dominance and Iterated BestResponse in Experimental '**p**-Beauty Contests'",American Economic Review, 88(4): 947-969.

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

Hoffmann, R. (1999): "The Independent Localisations of Interaction and Learning in the Repeated Prisoner's Dilemma", Theory and Decision 47: 57–72.

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

Hollis, M. and Sugden, R. (1993): "Rationality in Action", Mind, 102 (405): 1-35.

Hume, D. ([1739] 1978): *A Treatise of Human Nature*. Oxford: Oxford University Press.

Hurley, S. L. (1989): *Natural Reasons*, Oxford, Oxford University Press.

Hurley, S. (2003): "The limits of individualism are not the limits of rationality", Behavioral and Brain Sciences, 26:164-165.

Isoni A. Poulsen A., Sugden R., Tsutsui K., (2012): "Focal points in tacit bargaining problems: Experimental evidence", European Economic Review, Vol. 59:167‑188.

Janssen, M. C. V. (2001): "Rationalizing Focal Points", Theory and Decision, 50, 119-148.

Janssen, M. C. V. (2006): "On the strategic use of focal points in bargaining situations", Journal of Economic Psychology 27, 622–634.

Laffont J. (1975): “Macroeconomic constraints, economic efficiency and ethichs: an introduction to Kantian economics”, Economica, vol. 42, p. 430-437.

Levine D. (1998): “Modelling altruism and spitefulness in experiments”, Review of Economic Dynamics vol. 1, p. 593-622.

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

Marshall A. ([1890] 1920): *Principles of Economics*, London, Macmillan and Co.

Metha, J., Starmer, C. and Sugden, R. (1994): "The nature of salience: an experimental investigation”, American Economic Review, vol. 84(3): 658–73.

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

Nussbaum, M. (1986): *The fragility of goodness: luck and ethics in Greek tragedy and philosophy*, Cambridge University Press, New York.

Rabin M. (1993): “Incorporating fairness into game theory and econimics”, American Economic Review, vol. 83 p. 1281-1302.

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

Robles J. (2001): “Evolution in Finitely Repeated Coordination Games**”**, Games and Economic Behavior, 34(2): 312-330.

Saari D.G. (2002): “Mathematical social sciences; An Oxymoron?”, PIMS distinguished chair lecture. Pacific Institute for Mathematical Sciences.

Sally, D. (1995): "Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiment from 1958 to 1992," Rationality and Society, 7: 58-92.

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

Segal U., Sobel J. (1999): “Tit for Tat: Foundations of Preferences for Reciprocity in Strategic Settings” Mimeo. University of California at San Diego.

Smerilli A. (2012): “We-thinking and vacillation between frames: filling a gap in Bacharach’s theory”, Theory and decision, 73(4): 539-560.

Sober E., Wilson D. (1998): *Unto others: the evolution and psychology of unselfish behaviour*, Harvard University Press.

Stahl, D. O., Wilson P., (1994): “Experimental Evidence on Players’ Models of Other Players”, Journal of Economic Behavior and Organization, 25 (3): 309-327.

Stahl, D. O., Wilson, P., (1995): “On Players’ Models of Other Players: Theory and Experimental Evidence”, Games and Economic Behavior 10, 218–254.

Sugden, R. (1984): “Reciprocity: The Supply of Public Goods Through Voluntary Contributions”, The Economic Journal 94 (376): 772-787.

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

Sugden, R. (1995): “A theory of focal points”, Economic Journal, 105: 533-550.

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

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

Sugden R. (2004): *The economics of rights, cooperation and welfare*, second edition, Palgrave Macmillian, London.

Sugden, R.(2005): "Fellow-Feeling," in *Economics and Social Interactions*, ed. by B. Gui, and R. Sugden: Cambridge University Press.

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

Taylor M. (1976): *Anarchy & Cooperation*, John Wiley & Sons.

Taylor, M. (1987): *The possibility of cooperation*, Cambridge, Cambridge University Press.

Tsui K. and Weymark J. (1997): “Social welfare orderings for ratio-scale measurable utilities”, *Economic Theory*, 10: 241-256.

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

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

Vega-Redondo F. (1996): *Evolution, Games and Economic Behavior*, Oxford University Press, Oxford.

Vega-Redondo F. (2003): *Economics and the Theory of Games*, Cambridge University Press, Cambridge.

Weibull, J.W. (1995): *The evolutionary game theory*, Cambridge, Cambridge University Press.

Zizzo, D. J. (2004): “Positive Harmony Transformations and Equilibirum Selection in Two-Player Games,” Oxford: Department of Economics, University of Oxford.

Zizzo D.J., Tan, J.H.W. (2007): "Perceived Harmony, Similarity and Cooperation in 2 x 2 Games: An Experimental Study", Journal of Economic Psychology 28(3): 365-386.