



**HAL**  
open science

## Collective production processes, cooperation and incentives : experimental explorations

Benoît Chalvignac

► **To cite this version:**

Benoît Chalvignac. Collective production processes, cooperation and incentives : experimental explorations. Business administration. Université de Strasbourg, 2012. English. NNT : 2012STRAB007 . tel-00830965

**HAL Id: tel-00830965**

**<https://theses.hal.science/tel-00830965>**

Submitted on 6 Jun 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



*ÉCOLE DOCTORALE AUGUSTIN COURNOT*  
Bureau d'Économie Théorique et Appliquée

# THÈSE

présentée par

**Benoît CHALVIGNAC**

soutenue le 10 décembre 2012

pour obtenir le grade de

**Docteur de l'université de Strasbourg**  
en Sciences Économiques

**COLLECTIVE PRODUCTION PROCESSES,  
COOPERATION AND INCENTIVES  
EXPERIMENTAL EXPLORATIONS**

**THÈSE dirigée par :**

**M. LLERENA Patrick**  
**M. COHENDET Patrick**

Professeur, Université de Strasbourg  
Professeur, Université de Strasbourg

**RAPPORTEURS :**

**M. MARENGO Luigi**  
**M. MASCLET David**

Professeur, Scuola Superiore St Anna, Pise  
Directeur de recherches, Université de Rennes I

---

**SUFFRAGANTS :**

**M. GARROUSTE Pierre**  
**Mme ROZAN Anne**

Professeur, Université de Nice - Sophia Antipolis  
Professeur, ENGEES, Université de Strasbourg



*L'université n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses. Ces opinions doivent être considérées comme propres à leurs auteurs.*



## ACKNOWLEDGEMENTS

Writing this thesis has been very much like a journey in a land of which one has only an incomplete map, and to which one may never return if ever one manages to find some way out. This document is thus much like travel notes, where only parts of the experience are gathered and imperfectly reflected, as perceptions can only imperfectly be translated in words. The journey itself is the goal.

It is a journey into which I might not have embarked without the initial impetus from Patrick Cohendet, and that may have precociously come to an end, at times, without his support and that of Patrick Llerena. I also owe to my supervisors the freedom to explore, a scarce resource that I extensively needed, and am grateful for their patience. I am also thankful to Pierre Garrouste, Luigi Marengo, David Masclet and Anne Rozan, who have accepted to review this work.

Programming the experiments would have been a quite more challenging task without Kene. Paul was of invaluable help when designing the first one and collaborating with Herrade was a highly rewarding way to tackle the last one.

The leap from the final drafts of the chapters to this document has been eased by the cautious eyes of Jon, Natalia, Robin, Moritz and Monique. I thank you all for your helpfulness and kindness.

Jae Seung, Uyen and Stéphane welcomed me in Bureau 154 a few years ago, and were among the first ones to give me examples that one day, a thesis can be completed. In the remote territories of other doctoral offices or freshly promoted doctors, Yves, Julien, Guillaume, André, Paul, Claude, Sukrye, Taieb and a few others made me feel I was sharing a common destiny : other lands, other incomplete maps, same fate. Marie, Tricia, David, Li, Nathalie, Moritz, Mingfeng, René started around the same time as I did, and I am happy to join them in the post-doctoral club - sorry I'm a bit late, traffic on the road - and to be given the opportunity to thank them for their friendliness. Ani, Adeline, Agnès, Aleesha, Anta, Charlotte, Frédo, Jean-Philippe, Wallya and whole new generations, reminded me that time was up to hand in the annotated map - no time to wander any more, no time for more details.

During a discussion with Danielle, in February 2011, I dared to mention spring 2012 as an expected time of completion. She insisted that 2011 should be the year, no matter what. So I tried my best and here we are. Thank you Danielle.

Acknowledgements would not be complete without mentioning the daily support provided by close friends and relatives, but space is scarce and they already know. However, thank you Jean-Marc for your supportive "Allez Benoît!" on a fortuitous encounter around a year ago. It worked.

*La vérité se protège elle-même.*

*Les antagonismes croissent*

*autour d'elle avec*

*symétrie,*

*sans l'atteindre.<sup>1</sup>*

---

<sup>1</sup>*Le Jour et la Nuit : Cahiers de Georges Braque, 1917-1952*, Paris, Gallimard, 1952, p. 26.





# SHORT CONTENTS

GENERAL INTRODUCTION.....	13
PART ONE.....	15
CHAPTER I	
Economic theories of collective production and cooperation.....	17
1.Introduction.....	18
2.Enforcing cooperation: a matter of incentives.....	19
3.Mancur Olson and The Logic of Collective Action.....	25
4.Barnard, Simon, Cyert, March and the behavioral theory of the firm.....	28
5.Learning organizations: knowledge, routines and communities.....	33
6.The game theoretical analysis of social dilemma.....	35
7.Looking forward: Norms, institutions, habits and motivation.....	38
8.Conclusion: irreconcilable views?.....	43
CHAPTER II	
Illustrating the double edge of cooperation.....	45
1.Introduction.....	46
2.Symmetric two-person interactions: prisoner's dilemma experiments.....	48
3.Symmetric, n-person interactions: public good game and n-prisoner's dilemma experiments.....	51
4.Asymmetric two-person interactions: principal-agent settings, the 'gift-exchange game' and the ultimatum/dictator games.....	70
5.Asymmetric n-person interactions: three-player distribution games.....	79
6.Conclusions.....	81
PART TWO.....	85
CHAPTER III	
Voluntary participation in a collective-good game.....	87
1.Introduction.....	88
2.Related literature on voluntary participation.....	89
3.Experimental settings and treatments .....	91
4.Results and discussion .....	96
5.Comments .....	107
6.Conclusion.....	109
CHAPTER IV	
Alternative designs for cooperation .....	111

SHORT CONTENTS

1.Introduction.....	112
2.The experiment.....	115
3.Results.....	118
4.Discussion.....	128
5.Conclusion.....	129
CHAPTER V	
Disentangling distributional motives.....	131
1.Introduction.....	132
2.Related literature.....	132
3.A simple distribution game.....	137
4.Results.....	143
5.Discussion.....	146
6.Conclusion.....	148
GENERAL CONCLUSION.....	151
APPENDICES.....	155
Appendix A: Instructions to subjects - Chapter 3.....	155
Appendix B: Instructions to subjects - Chapter 4.....	159
Appendix C: Instructions to subjects - Chapter 5.....	164
REFERENCES.....	167
CONTENTS.....	179

---

## GENERAL INTRODUCTION

---

The study of knowledge creation processes has pointed to the complexity of individual interactions within productive organizations. This complexity appears to be such that incentive-based theories of the firm, which focus on information processing issues, may fail to grasp a substantial part of the individual decision-making involved in the context of organizational learning, and more broadly in collective production processes. Some of the questions raised by the knowledge-based theories of the firm do indeed have deeper implications than the specifics of knowledge creation and can be extended to most collective production settings. They had merely been ignored, for it requires different behavioral assumptions than those made by the standard contract theories to take them into account. Since collective production processes require a group of individual agents to contribute consistently in order to achieve a common goal, relevant theories must explain the necessary phenomenon of cooperation between these agents. While contract theories consider cooperation as deriving from the successful design of incentive mechanisms, knowledge-based and more generally behavioral theories of the firm emphasize the coordinating role of organizations where cooperation emerges from spontaneous contribution of the participating agents. Though both approaches might grasp a part of the picture, their contradicting behavioral assumptions must be revisited before any step to bring their diverging views into a unified framework can be taken.

Meanwhile, numerous experimental results have opened the narrow behavioral framework of monetary maximizing in settings where contract theories would have predicted failure in achieving the social optimum. These results have helped highlighting the importance of equity, peer pressure or more generally social norms and non-monetary concerns in explaining agents' behavior and, of equal importance, have given strong evidence of the possible emergence of cooperation in so-called non-cooperative situations.

The defining idea of this thesis is that the experimental exploration of the determinants of cooperation can help, at least to some extent, advancing and refining the behavioral assumptions on which theories of collective production are based. Our effort can be

considered as **an attempt to refine the behavioral assumptions of contract theories, in the light of experimental results, so that they would allow for the complex processes described by organization theories and the related resource-based view of the firm.**

The thesis is divided in two parts.

The **first part** is devoted to the theoretical and experimental literature reviews. **Chapter 1** examines the respective views of cooperation supported by the diverse theories of collective production processes, as well as some broader behavioral theories offering a framework in which the alternative families of theories of the firm could tentatively be brought together. **Chapter 2** presents a thorough review screening a large range of experimental results on the determinants of cooperation in various social dilemma situations, underlining the variety of individual choices observed and the strong complementarity between the alternative views of cooperative behavior.

In the **second part** of the thesis, we present three experiments in which we try to extend the research efforts presented in the experimental literature review, in line with our aim of refining the behavioral assumptions of contract theories. In **Chapter 3** we investigate the effect of voluntary participation on contribution levels in a public good game. It can indeed be argued that in real case situations agents most often face the choice to participate or not in the collective process at stake before choosing their contribution level, whereas the usual experimental settings is such that subjects do not have their say and must decide from within the social dilemma situation. **Chapter 4** gives an account of a gift-exchange game experiment where principals can decide to resort to an incentive device or not. We introduce a new treatment where each principal (agent) is paired with the same agent (principal) for all rounds, in order to measure the combined effect of repetition between fixed pairs and of the availability of both incentive and incentive-free mechanisms. We designed the experiment presented in **Chapter 5** in order to define the scope of distributional preferences, which might interfere with social outcome considerations in social dilemma experiments. At a more exploratory level, our settings also allowed us to propose an experimental test of diverging views on principles of social justice.

# Collective production processes, cooperation and incentives

---

---

## PART ONE

### *LITERATURE REVIEWS*

---

---

**CHAPTER I** : *Economic theories of collective production and cooperation*.....15

**CHAPTER II** : *Experimental explorations of individual behavior in social dilemma* .....43



---

---

## CHAPTER I

### ECONOMIC THEORIES OF COLLECTIVE PRODUCTION AND COOPERATION

---

#### Summary

In this chapter we give a broad overview of the various approaches of collective production in economic theory and of the corresponding conceptions of cooperation. Since these approaches originate from different traditions and sometimes from various fields they are far from constituting an integrative analysis of cooperation and of its behavioral drivers. However, through their diversity they provide a rich pool of analytical material supporting two opposing views of cooperation: (i) a behavior to be elicited and (ii) an emerging phenomenon.



## 1. Introduction

Cooperation can be considered to be at stake in any situation where the output of a process involving several human agents depends at least for some part on their joint action, meaning that this output cannot be totally decomposed back to the additive input of the concerned agents. Whenever the accomplishment of a given task is beyond the power of a single individual an underlying form of cooperation must be assumed if the outcome of the task is to be actually observed. Accordingly, cooperation is ubiquitous in any economic system and is clearly a fundamental reason why human beings have gathered in groups and societies and why those groups and societies have developed new productive abilities.

Having identified cooperation as a universal phenomenon we now have to track its interpretation in the terms in which economic analysis depicts human interactions, and more specifically, in the way economic analysis explains *how humans collectively manage to achieve goals that one cannot achieve alone*. Collective production processes have been studied by scholars in economics through the analysis of the main institutional forms in which they can be observed: firms, and more generally organizations. As we shall see in this chapter, alternative or complementary theories of the economic organizations reflect alternative, though maybe also complementary views of cooperation. Cooperation can be conceived as a behavior to be elicited from diverging interests or as an emergent property stemming from broader social interactions among agents. Though experimental studies are limited to strictly monetary settings for methodological reasons, and are thus better suited for testing theories that rely on the former conception of cooperation, they also illuminate behavioral patterns which support the latter one, even in this narrower context.

Given the density of literature in the field of the theory of the firm, we purposely limited the scope of the present review to the most representative work and only included what we considered as necessary to highlight the behavioral assumptions at stake. To our knowledge, most of the many rich developments posterior to the mentioned references relied on the same assumptions or did not focus on the individual choices that determine the elicitation or the emergence of cooperation.

An important contrast can be drawn between contract theories (section 2), which hold that cooperation can only occur through the setting of narrow incentive schemes, and behavioral and knowledge-based theories (sections 4 and 5 respectively) which focus on the

convergence of interest between the organization and its individual members. Starting from the same assumptions as the contract theories, Mancur Olson (section 3) proposed an intermediate way for cooperation which does not require the formal contractual setting of incentives but still relies on the self-interested choices of agents. Similarly, recent developments in game theory (section 6) integrate various complementary factors in the utility function of the players which allow for the emergence of some form of cooperation in incentive-free settings. The analysis of the role of norms, institutions, habits, and intrinsic motivation (section 7) in the determination of individual behavior, though pertaining to sometimes unrelated original streams, provide promising sources for future integration of these contrasting views of cooperation.

## 2. Enforcing cooperation: a matter of incentives

### 2.1. *A familiar picture: the selfish agent*

"But man has almost constant occasion for the help of his brethren, and it is in vain for him to expect it from their benevolence only. He will be more likely to prevail if he can interest their self-love in his favour, and shew them that it is for their own advantage to do for him what he requires of them." (Smith, 1902, p.56)

"What is now commonly understood by the term 'Political Economy' [...] does not treat of the whole of man's nature as modified by the social state, nor of the whole conduct of man in society. It is concerned with him solely as a being who desires to possess wealth [...]. It predicts only such of the phenomena of the social state as take place in consequence of the pursuit of wealth. It makes entire abstraction of every other human passion or motive; except those which may be regarded as perpetually antagonizing principles to the desire of wealth, namely, aversion to labour, and desire of the present enjoyment of costly indulgences." (Mill, 1844, p.137)

Though both Adam Smith and John Stuart Mill used careful methodological reserves for their descriptions of human behavior, as embodied by the then-emerging economic theory, their conceptions paved the way for the characterization of the selfish, short-sighted agent on which the contract theories built their models of collective processes and cooperation. The original dichotomy operated by Mill singling the selfish desire for wealth out among "the variety of desires and aversions which are concurrently operating upon [man]" (Mill, 1844,

p.139) led to the mere ignorance of the latter in the subsequent analysis which we will now address.

## 2.2. *Property rights*

Assuming the existence of a source of gain from cooperative activity, Alchian and Demsetz (1972) ask: "How can members of a team be rewarded and *induced* to work efficiently?" The implicit assumption on individual behavior is quite straightforward, agents will engage in cooperative actions only to the extent that a contractual arrangement can be found that shapes the incentive structure compatibly with some desirable outcome. This implies, first, that the first-best outcome - when agents work efficiently in the collective production process without bearing the cost of the incentive structure - is ruled out. This also implies that agents have *ex ante* a clear perception of the potential benefits of cooperation and that some kind of invisible hand or driving force is at work that will fit and gather the concerned stakeholder of the organizational entity in a seemingly spontaneous ordering. We will discuss further in this thesis the latter implication and concentrate now on the former one.

As surprising as it may appear without a word of explanation, shirking comes into the picture as soon as individual action is not fully observable without a cost. The justification is to be found in the existence of "a person's utility function" in which "both leisure and higher income" enter. The private cost of shirking being lower than the true total cost to the team, the individual trade-off that maximizes the utility function will be socially suboptimal as long as asset ownership will be shared between the members of the team.

## 2.3. *Agency theory*

Jensen and Meckling (1976) define the agency relationship as "a contract under which one or more persons (the principal(s)) engage another person (the agent) to perform some service on their behalf which involves delegating some decision making authority to the agent". Accordingly, the agency problem is "the problem of *inducing* an agent to behave as if he was maximizing the principal's welfare". This problem is "quite general" and "exists in all organizations and in all cooperative efforts". It seems to follow that cooperation is the sole result of the use by a principal of the available means at his disposal directed at driving an agent's behavior. Then any economic relationship between two actors of a production process cannot be but asymmetric, hierarchical and based on the instrumental animation or manipulation of one by the other. The cooperative outcome denotes a skillful behavior of a

principal who managed to overcome the agency problem and achieved to turn the selfish agent into a promoter of his own selfish interest. Cooperation is thus not the realization of a common interest but a by-product of an asymmetric interaction triggered by the principal for the realization of his individual interest. Though it seems hard to measure the accuracy of this conception one can reasonably argue that it may reflect a significant part of all economic interactions. The behavioral assumptions from which stems the identified problem still needs to be stated: "If both parties to the [agency] relationship are utility maximizers, there is good reason to believe that the agent will not always act in the best interests of the principal. The principal can limit divergences from his interest by establishing appropriate incentives for the agent and by incurring monitoring costs designed to limit the *aberrant* activities of the agent" (emphasis added). Here is the clue, and Jensen and Meckling strongly insist on the point: stating that "[they] retain the notion of maximizing behavior on the part of all individuals in the analysis that follows", they refer in the following footnote to "the fundamental importance of the assumption of resourceful, evaluative, maximizing behavior on the part of individuals in the development of theory" which is advocated by Meckling (1976). Though their paper is explicitly directed to the specific analysis of "contractual arrangements between the owners and top management of the corporation" (p.7) their claim is that these behavioral assumptions are sound enough to found a general theory of organizations.

A typical trait of this stream of analysis, an *ex ante* division of production factor endowments, capital and labor, is assumed here. The question of cooperation is thus restricted to the correct alignment of incentives among exogenously endowed individuals, and the dynamic interactions between the organization as a whole and each of its members are ignored. In a general cooperative situation, however, every actor can be simultaneously considered to be both the principal and the agent to every other actor.

#### **2.4. *Multiple agents: teams***

Holmstrom (1982) tackles "the problem of free riding when there is joint production" and shows that "non-cooperative behavior will always yield an inefficient outcome if joint output is fully shared among the agents". Though he rejects the focus put by Alchian and Demsetz (1972) on the monitoring role of the principal, claiming instead that "the principal's primary role is to break the budget-balancing constraint" or "to administer incentive schemes that *police agents* in a credible way", the behavioral assumptions remain unchanged: agents have a preference function (which they are implicitly assumed to maximize) increasing in money

(their share of the joint monetary outcome) and decreasing in the private non-monetary cost of effort ("non-observable action"). In other words, here again, only incentives that oppose the force of pursuing one's individual, narrow and myopic interest will have some chance of generating sustainable collective production. We qualify the individual interest described here as "narrow" for it does not take into account the individual benefit that could be rationally expected from the collective interest at stake (the fact that the collective interest is actually included in the individual one is ignored), and as "myopic" in the sense that only sequentially independent courses of action, or choices, are considered in the deliberation.

Interestingly enough, it appears that in both Alchian and Demsetz (1972) and Holmstrom (1982) the principal is supposed to be created or brought in by the group or team of agents in the need of a solution to the free-rider problem. Though gathering empirical data and historical observations was somehow unfashionable in mainstream economic theory at the time of the publication of these two papers, it is still quite surprising that they came so easily, and without more scrutiny, to the idea that the origin of the capitalistic firm<sup>2</sup> is to be found here. Economic history tells a rather different story of landowners and successful merchants who gathered and hired workforce in order to leverage work-intensive entrepreneurial opportunities, which required large amount of capital investment and were mostly driven by technological change<sup>3</sup>.

Still, applying the above mentioned behavioral assumptions to the public-good-game-like situation of team production, it follows that "there do not exist sharing rules"<sup>4</sup> such that the highest possible joint monetary outcome is produced and fully allocated among the agents. "[This] result", Holstrom goes on, "indicates that in closed (budget-balanced) organizations like a labor-managed firm or a partnership, free-rider problems are likely to lead to an insufficient supply of productive inputs like effort. This observation is the starting point for Alchian and Demsetz' (1972) well-known theory of the firm. They argue that the inefficiency of a partnership will cause an organizational change. To secure a sufficient supply of effort, firms should hire a principal to monitor the behavior of agents. The monitor should be given title to the net earnings of the firm so that he has the proper incentives to work. Such an arrangement will restore efficiency. At the same time, it will change the partnership into a capitalistic firm with the monitor acting effectively as the owner."<sup>5</sup> Fully subscribing to the

---

<sup>2</sup> The phrase 'capitalistic firm' is used in the literature for contractual arrangements where ownership is not held by the employees, as opposed to 'cooperative firms'.

<sup>3</sup> On this point, see e.g. Polanyi (1957).

<sup>4</sup> Holstrom (1982), p.326.

<sup>5</sup> Ibid., p.327.

case made by Alchian and Demsetz, Holmstrom follows their argument concerning the most accurate form of productive organizations and similarly rejects the sustainability of "budget-balanced" organizations (cooperative or labor-managed firms) as a possible candidate for collective production. Indeed, as a consequence of the free-rider problem, incentives must be introduced in the sharing rules in the form of penalties or bonuses, which can not be credibly enforced by the agents themselves so that "the enforcement problem can be overcome only by bringing in a principal (or a party) who will assume the residual of the non-budget-balancing sharing rules."<sup>6</sup> In this view cooperation must be induced by incentive schemes, the credibility of which lies in the existence of an enforcing entity, namely the principal.

Holmstrom (1982) nonetheless distances himself from "Alchian and Demsetz' story" in that he does not consider monitoring itself but the possibility to finance and implement incentive schemes as the key factor that makes the introduction of a principal a necessary condition for effective collective production: "there is little to suggest that either of the two forms of organization would stand at a comparative advantage when it comes to monitoring alone".

The conclusion concerning the comparative advantage of the capitalistic form of collective production holds anyway, and can also be found in the same paper in the following, reversed, formulation: "the fact that capitalistic firms feature separation of ownership and labor implies that the free-rider problem is less pronounced in such firms than in closed organizations like partnerships." Reformulated this way it somehow sounds either like a truism or a sophism. If the free-rider problem can be overcome by separating ownership and labor, then indeed firms where such a separation occurs will overcome that problem while the others might not. However, the stated implication holds only as long as the main argument of the paper holds: the fact that capitalistic firms feature separation of ownership and labor does not tell much about the comparative intensity of the free-rider problem.

This formulation furthermore seems to assume that the forces driven by diverging individual interests are not opposed or counteracted by incentive schemes but rather dissolved or removed in some way. This sharply contrasts with the behavioral assumptions that led to the conclusion that the Pareto-optimal outcome could not be reached by budget-balanced sharing rules. It is a fundamental law of physics that forces remain unchanged, whether opposed or not, as long as the conditions that create them remain, so that there is no reason to assume that setting an incentive scheme would mitigate the very behavioral drivers that it is made to

---

<sup>6</sup> A possible alternative to bringing in a principal is evoked : "In a dynamic context the punishment in [the corresponding equation] can be interpreted as a threat to discontinue cooperation" but not elaborated further.

oppose. With this in mind, one could postulate that where 'constrained' cooperation is achieved a degree of tension will be created that works against the collective production processes, either in the form of unnoticed shirking, slacking or even resource diverting, or in any of the multiple ways psychological forces tend to use to achieve the fulfillment of individual aspirations.

According to this stream of analysis, emerging cooperation is thus a utopia. Having ruled out the mere possibility of cooperation as a free choice to act in the common benefit of all parties, agency theories are not of much help once the studied processes do not mostly occur in situations where defective behavior is anticipated by at least one of the actors.

### 2.5. *Transaction cost economics (TCE)*

First, claim Coase and Williamson, there were markets. They acknowledge however that another institutional type, the economic organization labeled the 'firm', also processes some of the productive activities though it is a locus of relaxed competitive tension. In order to explain the existence of organizational forms in contrast with market mechanisms, it was then necessary to isolate properties that could provide some possible competitive advantage to the organization. According to Williamson, frictions or transaction costs occur and differ in some situations of the economic system due to a combination of human, technological and environmental factors.

More specifically, Williamson gave a central place in his analysis of economic institutions to bounded rationality<sup>7</sup> and opportunism. As summed up in Hodgson (2004a), "Williamson (1975) famously defined 'opportunism' as 'self-interest seeking with guile'. He argued that 'economic man is a much more subtle and devious creature than the usual self-interest seeking assumption reveals' (*ibid.*). Williamson (1985) later elaborated the concept of opportunism in terms of 'the incomplete or distorted disclosure of information, especially to calculated efforts to mislead, distort, disguise, obfuscate, or otherwise confuse'". Williamson then singles out the perception of the risk to face opportunistic behavior from some agents and identifies it as the only necessary condition for market contracting not to be the only way economic transactions are carried out. Much in the same flavor as property rights and principal-agent theories, TCE thus presents cooperation in collective production processes as the product of the successful binding of individual interests, augmented this time with guileful behavior.

---

<sup>7</sup> Williamson, 1991.

However, all the contractual forms that depart from the classical market contract also rely on a minimum amount of trust between the parties, which seems to contradict the fundamental assumption on opportunism. The hierarchical nature of the work contract, for instance, does not make it less prone to opportunistic behavior than the market contract. Moreover, like the other contract theories, TCE focuses on transactions and sets aside the complexity of the production process hosted by the organization. While TCE depicts the individual agent as busy with processing information and dealing with uncertainty stemming from information uncertainty, the daily organizational hurdles are ignored.

Thus the contract theory approaches describe the collective production processes as the sum of individual outputs exchanged through a series of transactions. Successful production is achieved through the successful overcoming of information asymmetries between agents that seek to exploit them in order to maximize their short-term utility, namely by securing the highest possible monetary income while minimizing their contribution. Similar behavioral assumptions will however lead Mancur Olson to slightly different conclusions.

### 3. **Mancur Olson and *The Logic of Collective Action***

"The corporation is expected to further the interests of its stockholders, that is, its members. This study does not follow the terminological usage of those organization theorists who describe employees as "members" of the organization for which they work." (Olson, 1971, p.6)

It is a significant paradox that the above quotation is to be found in one of the most prominent seminal works analyzing the intricate relationship between individual interests and collective action, as a footnote of Chapter I "Theory of Groups and Organizations". According to this definition of the corporation, only cooperative firms would fully qualify as an object for the study of cooperation, since their specificity is precisely that the employees are also the shareholders. Though Olson's analysis of collective action largely applies to collective production processes, whatever the contractual relationships at stake, his note on the "members" of the corporation draws the attention on the diversity of collective goals throughout the economic system, possibly overlapping within its constitutive organizations, and underlines the behavioral ambivalence that can follow.



However, the purpose of organizations, as it is indeed clearly made explicit on the next page, leaves no doubt about the subject covered:

"When a number of individuals have a common or collective interest – when they share a common purpose or objective – individual, unorganised action will either not be able to advance that common interest at all or will not be able to advance that interest adequately. Organizations can therefore perform a function when there are common or group interests, and though organizations often also serve purely personal, individual interests, their characteristic and primary function is to advance the common interest of groups of individuals." (Olson, 1971, p.7)

A group is then defined as "a number of individuals with common interests", which the organization is "*expected* to further". In other words, the organization is created with the purpose of pursuing goals that the individuals cannot achieve alone. The success of the organization in turn depends on the individual contributions of its members.

### ***3.1. Group size and individual weights***

The main argument of *The Logic* lies in the effect of group size on the incentive structure of the group. Assuming the "widely accepted premise of rational, self-interested behavior" of the individual members of a group, Olson opposes the view that this assumption is enough to claim that groups will in turn act in support of their collective interests. Self-interested individuals, he argues, will simply balance perceived benefits and costs of their contribution to the collective good. This balance is largely influenced by group size, because the extent to which contribution or non-contribution of one group member is noticeable by the other members depends on the total number of group members. When the group is small enough, strategic interaction between its members makes the provision of the collective good possible since the contribution choice by one member influences costs or benefits to the other members and thus their contribution choices. If enough members of the group take the effect of their contribution choice on other members' choice into account the collective good might be provided. Small groups can thus achieve some level of cooperation without coercive mechanisms or externally designed incentive schemes, and cooperation can emerge from the furtherance of a common interest by self-interested individuals.

Conversely, "latent" groups, defined as groups large enough so that the contribution decision of one member goes unnoticed by other members, are not expected to provide collective goods to their members without setting coercive or "separate, outside incentives".

### 3.2. *Inclusive vs. exclusive groups*

In a distinction that would now be expressed along the line of excludable vs. non-excludable and rival vs. non-rival goods, Olson differentiates between exclusive groups and inclusive groups. The latter are related to situations where some level of non-rival, non-excludable ("public") good is to be produced, while the former are concerned with the production of rival, non-excludable ("common") goods<sup>8</sup>. In the same note, Olson also refers to Buchanan's "club" goods (Buchanan, 1965), which are non-rival to some extent and excludable, as a case for the action of exclusive groups. Though common and club goods have seemingly opposite properties, both are crucially affected by the number of people who consume the good, club goods being at risk of congestion above a given threshold and common goods being exposed to over-consumption and depletion. Exclusive groups are thus under greater tension to control the size of their membership and to insure wide contribution from their members. Though this distinction is not central to his analysis, Olson underlines that "strategic interaction is less common and less important" in inclusive groups since non-rivalry lets the consumption of any group member remain unaffected by the consumption of other or new members. Consequently, group oriented action is more likely to be observed in inclusive groups than in exclusive groups, provided that the group is small enough to make contribution choices noticeable by other members.

### 3.3. *Social incentives and the mobilization of latent groups*

Most of the analysis of the effect of group size on collective action is implicitly based on monetary or, in the wording of Olson, "economic" incentives. Olson however does not pretend that other forms of incentives are absent from individual behavioral drivers. He also acknowledges the role of "social incentives" as rational motivations, and defines them as "the desire to win prestige, respect, friendship, and other social and psychological objectives". In cases where economic, or monetary incentives are absent, individuals might still contribute to the provision of the collective good if social incentives are present. More precisely, the economic gain that would arise from non-contribution would be outweighed by the social

---

<sup>8</sup> Olson originally referred to Head's (1962) distinction between "infeasibility of exclusion" and "jointness of supply".

loss incurred. Referring to the work of Barnard (1938 and 1948), Clark and Wilson (1961) and Simon (1957), Olson stresses that "social incentives must be analyzed in much the same way as monetary incentives" and thus largely reduces their influence to small groups, where members can interact with each other in a direct, personal way. He still identifies a case where social incentives may play a role in latent groups. This is if the large group is itself composed of a number of small groups – the "federal" group. In a very similar way, he acknowledges other types of incentives – "erotic incentives, psychological incentives, moral incentives and so on" – only to suggest that they have no different role than social incentives and thus fall short of mobilizing latent groups, except for the federal group and only through the incentive systems within the small groups forming the federation.

Though peer pressure, under the label "social incentives" is fully part of Olson's analysis, any other kind of behavioral drivers appear only in a footnote in the same section<sup>9</sup>. His conclusions on the provision of collective goods thus solely rely on the specific properties derived from the size of small groups, namely the personal strategic interaction between their members.

Olson should nonetheless be given credit for pointing to the possible success of monetary incentives in achieving a collective output without resorting to a central authority in charge of managing these incentives through contractual tools. His assumptions are however not fundamentally different to the assumptions outlined in contract theories, especially in that he considers the achievement of the collective output as the outcome of one-off contribution decisions by self-interested individuals balancing benefits and costs. The behavioral theories of the firm that we will review now start from the opposite side, taking the organization as given and considering the default choice of its members as continued participation.

#### **4. Barnard, Simon, Cyert, March and the behavioral theory of the firm**

##### ***4.1. Barnard: moral behavior and informal organization***

In his seminal contribution to the behavioral theory of the firm, Barnard (1938) attempts to describe and advance a theory of cooperative behavior in the context of formal organizations.<sup>10</sup> Observing that formal organization involves "conscious, deliberate, and

<sup>9</sup> Olson (1971), p.61, note 17.

<sup>10</sup> Barnard's work has recently been thoroughly reviewed and analyzed in Gabor and Mahoney (2010), from which we drew most of Barnard's quotations used in this section.

purposeful cooperation among people"<sup>11</sup>, he stresses the promotion of communication among these people as a fundamental function of organizations. In his picture, the organization also acts dynamically in order to preserve some coherence between organizational goals and the individual goals of its various stakeholders. The role of incentives in formal organization is fundamental. Inadequate incentive designs lead to the dissolution of the organization, to the distortion of organizational goals, or to failure of cooperation. He insists on the tendency of formal organizations to undergo failures: "failure to cooperate, failure of cooperation, failure of organization, disorganization, dis-integration, destruction of organization—and reorganization—are the characteristic facts of human history."<sup>12</sup> Formal organizations cannot endure without maintaining some level of informal organization, which allow for the expansion of cooperation in spite of the deleterious effect of formal organizations on personality. In his view, this has much broader implications than the fake of organizational entities, since "the necessary expansion of cooperation and the development of the individual are mutually dependent realities and a due proportion or balance between them is a necessary condition of human welfare"<sup>13</sup>.

Within organizations anyway, moral behavior is indispensable to the survival of formal organizations and cannot be reduced to self-interest motives: "I mean by moral behavior that which is governed by beliefs or feelings of what is right or wrong *regardless of self-interest* or immediate consequences of a decision to do or not do specific things under particular conditions"<sup>14</sup> (emphasis added). No collective production process could possibly be pursued without some level of confidence concerning the integrity, honesty and sincerity of the management. Even though formal organizations and the provision of adequate incentive structures are presented in Barnard's work as necessary conditions of the achievement of common goals, "organizations endure, however, in proportion to the breadth of morality by which they are governed. This is only to say that foresight, long purposes, high ideals are the basis for the persistence of cooperation"<sup>15</sup>. Formal organizations that can be supported by informal organization where members transcend short-term interest will thus have a stronger position than organizations where immediate, material incentives prevail.

From his experience as an executive, Barnard is further able to give vivid examples of daily behavior within organizations, suggesting that "if you will stop taking the business man at

---

<sup>11</sup> Mahoney (2005).

<sup>12</sup> Barnard (1938), p.5.

<sup>13</sup> Id. p.296.

<sup>14</sup> Barnard (1958), p.4.

<sup>15</sup> Barnard (1938), p.282.

his word and quietly watch him when he is off guard, you will find he is taking care of poor old John who couldn't be placed anywhere else, that he is risking both profit and failure rather than cut wages, that he continues an unprofitable venture on nothing but hope rather than throw his men out of work. Much of this is unsound. It would be better if economic motives did operate more effectively, but the point is that it is impossible to get to the root of personnel relations or understand labor troubles or successes on *the unrealistic assumption that economic motives exclusively govern. They merely limit and guide.* They control more in some cases or some businesses than others"<sup>16</sup> (emphasis added). In his affirmation of spontaneous cooperative effort he also states that he has himself "seen large groups of employees voluntarily and wholeheartedly cooperate to increase individual and collective efficiency and production in order to reduce expenses when it was recognized that the immediate effect was to the pecuniary disadvantage of the employees themselves. The importance of such collaboration to all involved is incalculable. It is neither justified, nor can it be obtained, except on the basis of a confidence inspired by experience"<sup>17</sup>.

Barnard's analysis thus dramatically expands the scope of behavioral drivers, drawing a much more contrasted picture where the role of material incentives and self-interest is tempered by moral behavior and some sense of a collective interest. This clears the way for an analysis of the collective production processes focusing on coordination and learning.

#### **4.2. *Simon: bounded rationality and limited attention***

Simon (1941) opens with the observation that "administration is most usually discussed as the art of getting things done", by which "principles are set forth for securing concerted action from groups of men." Simon thus follows Barnard (1938) in considering organizations as systems of cooperative behavior. "Principles of administration abstract from the particular objective of any specific organization, and, if valid, are applicable generally to any group of men organized for purposive action". His main focus will however be on the decision processes involved and, in particular, on the boundedness of rationality at stake within organizations.

Contrasting "neoclassical economic man" which "maximizes - selects the best alternative from among all those available to him" with "organizational man" which "satisfices - looks for a course of action that is satisfactory or good enough", Simon (1997) clarifies further his conception of rationality writing that "behavior is termed 'rational' when it is appropriate to

---

<sup>16</sup> Barnard (1948), p.15.

<sup>17</sup> Barnard (1948), p.12.

the desired ends"<sup>18</sup>. It follows that cooperation will not be achieved as a consequence of the successful alignment of diverging interests so much as through the coordination of choices between alternative means to reach the collective goals.

Regarding the participation of the members to the organization, the choice to participate or not can be viewed as pertaining to the individual ends, following a valuation process which Simon clearly states as being intrinsically non-rational: "the system of values is in itself a non-rational construct. [...] Propositions defining the value-system (ethical propositions in the pure sense) involve no factual elements. They are indicative only to the extent that they say something about the psychology of the subject himself"<sup>19</sup>. The realization of collective goals, through collective production processes, thus critically depends on the ability of the organization to deal with, and ultimately reduce, the complexity of its environment to allow 'satisficing' choices to be identified and selected in order to pursue organizational goals, while providing its members with satisfying inducements. As long as individuals perceive the organizational goals and their own contribution thereto as supporting their personal ends – as appropriate means to their personal ends – they keep contributing, allowing the organization to survive and fulfill its aims as a result of the individual contributions. The content of personal goals, and thus the nature of the inducements provided by the organization, is however left open since the valuation of ends is out of the realm of rational deliberation.

Simon (1997) also importantly introduced learning as a major component into the analysis of collective production processes. Since the organization must learn in order to economize on its cognitive resources, because of limited attention, it must also memorize from its experience to be able to reproduce behavior that has proved successful and efficient. At the individual level, habit is an important mechanism that supports the preservation and replication of useful behavioral patterns. At the level of the organization, which has to set an artificial system of memory in the form of repertoires of collective behaviors, organizational routine is the repository of behavioral patterns. Organizational learning thus implies the modification of routines as a result of a trial and error, problem solving process. The importance of problem solving and learning for organizations, together with the complexity of the processes involved, implies that the choice to contribute must be closely tied to the choice to participate or not, the two being linked by the acceptance of the authority relationship, and that participation must be a stable, long term decision. Cooperation thus

---

<sup>18</sup> Simon (1947), p.32.

<sup>19</sup> Simon (1941), p.75.

emerges through the coordination of individual behavior and is self-sustained as long as the achievement of organizational goals provides the necessary inducements for the members of the organization to maintain their participation.

#### **4.3. *Cyert and March: subcoalitions and bargaining***

Cyert and March (1963) are concerned with the business organization, and more specifically the decision-making process in the firm. Examining the internal operations of the firm they focus on organizational goals, expectations, choice and control.

Considering the organization as a coalition of stakeholders, some of them being organized into subcoalitions, Cyert and March describe how organizational goals are set as the result of an ongoing bargaining process between stakeholders. Setting budget and allocating tasks are tools for mutual control, allowing collective decisions to be made in spite of potentially conflicting goals among stakeholders and subcoalitions. Importantly, Cyert and March underline that unresolved conflicts between stakeholder goals is an ongoing feature of organizations. This is due to the sequentiality of the bargaining process stemming from limited attention, and to the necessity to maintain standard operating procedures unless they fail to meet the selected goals. Thus, conflicting individual interests need not be aligned, or only partially, to observe the cooperative output produced by the organization.

Furthermore, the account of organizational learning given by Cyert and March highlights the intensity of cognitive activity that business organizations usually require from their members. Since the monetary compensations offered to the latter is most often determined on the basis of a "business as usual" description of the concerned position, it more closely resembles a show-up fee rather than proper compensation for either the entire effort provided or for the actual value created by the individual contribution to the process.

The dynamics of the organization as a bundle of subcoalitions described in Cyert and March (1963) somehow echoes Olson's (1971) conclusions on the role of social incentives in the mobilization of large groups, when composed of a number of small groups.

#### **4.4. *The "behavioral" agent***

As pictured by the behavioral theory of the firm the human agent, or "organizational man" in Simon's words, is engaged in a succession of choices most of which consist in or lead to the repetition of previous actions. Search is engaged only to the extent that a new problem arises in the form of an observed inappropriateness of known or habitual means to present ends,

and the intensity of search is increasing with the acuteness of the problem at stake. Hence, the questions arising from the apparent tension between self-interested choices and the achievement of collective goals is lifted while focus shifts from choice motivations to choice processes and the actual content of rational decision-making. As will also be the case in evolutionary and knowledge based views of the firm, though maybe in a more implicit flavor, collective production is achieved and sustained through the enactment and the repetition of collective processes that cannot be reduced to a sum of individual actions. The behavioral assumptions are not focused on short-term decisions aimed at maximizing monetary gains but on the "organizational equilibrium".

The organizational equilibrium is a central concept to both Barnard and Simon, which mostly implies a sustained participation of the agent and in turn the provision by the organization of inducements equal or higher than the satisfactory level set by the agent. Focusing on long-term commitment and based on trust relationships between the organization and its members, this view clears the way for the analysis of organizational dynamics and for the study of organizational learning processes.

## **5. Learning organizations: knowledge, routines and communities**

Building on Simon, Cyert and March and bringing forth Schumpeter's dynamic vision of economic systems and institutions, the evolutionary theory of the firm (Nelson and Winter, 1982, Teece et al. 1994) is particularly focused on the importance of the collective production processes as driven and coordinated by routine, and on the knowledge creation process resulting in the evolution of routines. Through successive mergers and contributions from management science, a resource-based or knowledge-based theory of the firm emerged which draws on the complexity of learning processes to emphasize the major role of coordination and cognitive activity in collective production processes.

As in Simon (1997) and in Cyert and March (1963), the focus here has completely shifted from aligning diverging interests to coordinating the operational processes and the search activity of the members of the organization through mobilizing and enacting the knowledge encapsulated in the firm as a "community of communities". The learning process has indeed been shown to be dependent on various types of communities<sup>20</sup> and on repeated, on-going

---

<sup>20</sup> See Brown and Duguid (1991), Cohendet and Llerena (2003).



and multi-directional social interactions<sup>21</sup>. In this view, the activities of the business firm must be hierarchically classified<sup>22</sup>, the core competencies stemming from the idiosyncrasies of organizational learning needing to be managed according to the specificities of knowledge creation processes and the periphery according to the make-or-buy decision. In line with the conception of routine as "truce among conflicting interests"<sup>23</sup>, which initially came without much elaboration, the preliminary decision to participate in the collective process is however left open concerning the domains of the core competencies. In the latter, the critical role of communities further enlightens the importance of non-material interests (reputation-building or intrinsic motivation) as well as long-term expectations and trust in collective production processes.

In a preliminary attempt to bridge incentive-based and competence-based views of the collective production institutions, Dosi et al. (2003) assume "a weak incentive compatibility to begin with (see Dosi and Marengo, 1994) in the loosest sense that there exists some selection pressure which, in turn, generates some connection between performance and rewards." As a choice must be made about which side the first pier of the bridge will be laid down, "however, having that, one precisely focuses (as a first theoretical approximation) on the diverse problem-solving characteristics of different organizations, and only in the second instance one tackles the ways in which incentive structures interact with problem-solving knowledge." In other words, the vision of the organization as a locus of knowledge creation or problem solving assumes that some preliminary, sufficient level of incentives is set so that members maintain their participation choice and get into the collective process.

Though individual interests are at stake in the case of cognitive communities, either in the form of reputation effect or as a trade-off between building trust relationships and choosing the hit-and-run option, the complexity of the individual motivations in this case is such that they cannot be reduced to short-term monetary trade-offs and are thus beyond the scope of the experimental explorations presented in the following chapters of this thesis. However, the behavioral patterns supporting the emergence of cooperation in the context of monetary stakes open the way in providing a basis for the study and explanation of more complex collective production processes. We will now move to the analysis of individual choices in such settings provided by a natural candidate - game theory.

---

<sup>21</sup> See Nonaka and Takeuchi (1995).

<sup>22</sup> As argued in Cohendet and Llerena (2003).

<sup>23</sup> See Nelson and Winter (1982).

## 6. The game theoretical analysis of social dilemma

The nature of collective production processes, when considered from the point of view of the individual agent, seems to lend itself to game-theoretical analysis. Game theory can indeed be regarded as a multi-agent decision theory analyzing the decision-making process when there is more than one decision-maker and where each agent's payoff possibly depends on the actions taken by the other agents.

Each player is importantly assumed to behave rationally. Rationality, in the language of game theory, implies that each player tries to maximize her payoff irrespective to what other players are choosing or getting. In essence, each player is in a situation to decide a set of moves which are made available according to the rules of the game and which maximize her rewards.

Because of the generality and abstraction of the games formulation, payoffs have traditionally been expressed without reference to specific, individual preferences and as independent from the players involved. In order to derive tractable solutions, game theory cannot indeed allow for much heterogeneity and variability in players preferences.

### 6.1. *Nash vs. Pareto optimality*

In its narrowest definition, a stage game solution is a Nash equilibrium if no player would be better off by choosing a different strategy unilaterally. If cooperation situations are simplified enough to be simultaneous, context independent, one-shot non-cooperative games, which usually implies that pay-offs are expressed in monetary terms, the Nash equilibrium happens to be the non-cooperative solution.

It can easily be argued that collective production processes often imply that at least some of the participants put aside immediate monetary gains that they could realize through short-term action, or relinquish some alternative project of their own that might yield higher pay-offs. The assumptions leading to the failure of cooperation thus needed to be loosened.

### 6.2. *The evolutionary game theory approach: reciprocity and punishment*

Since cooperation appears as a strong factor in regards to the survival and development of the human species and since it is widely observed in situations where it provides no immediate gain to agents, some selection mechanism has been assumed to support its

adoption and its emergence as a dominant, evolutionary stable, strategy among actors that are making decisions based solely on their own material situation only.

The repetition of social interactions, between non kin as well as between kins, has been shown to support cooperative behavior in bilateral relationships, as argued for instance in Axelrod and Hamilton (1981) and in Trivers (1971). The short-term gains induced by the non-cooperative choice of the agent would indeed be offset by the prospective losses associated with the non-cooperative choice of the other agent in future interactions. This relies on the assumption that reciprocity, aka the 'tit-for-tat' strategy, will determine the behavior of the other agent, who will choose to cooperate if the first partner has cooperated and not to cooperate otherwise. As in Friedman (1971) or Fudenberg and Maskin (1986), the latter choice can be interpreted as a punishment targeted against the non-cooperative choice, which in turn creates an explicit or an implicit credible threat and possibly deters defection even within a population a defection-prone individuals.

Building on the analysis presented in Gintis (2000) and subsequently developed in Bowles and Gintis (2004), and departing from the above arguments which rely solely on materially interested agents, Fehr et al. (2002) make a case for 'strong reciprocity', by which they denote the tendency of "many people to voluntarily cooperate" and to punish defectors at their own cost. Accordingly, 'strong reciprocators', a term which may be more accurately defined as agents who adopt a strong reciprocity strategy in a given situation, are willing to sacrifice resources to reward cooperative behavior or punish non-cooperative behavior, even without the perspective of a foreseeable return. This stream of literature is closely related to work on social norms, since these are considered to be the behavioral drivers that when enacted gave small groups a competitive advantage, and explain the sustainability of the cooperative choice even when it came at a net material loss for the individual.

### ***6.3. Taking the utility function one step beyond: fairness and equity***

Drawing on the framework developed by Geanakoplos and al. (1989), which introduced players beliefs as determinants of their payoffs in addition to their actions, Rabin (1993) defined 'equitable payoffs' from which the player's 'kindness function' can be derived. The player's utility can now be defined by incorporating both the material payoffs and the perceived fairness of each outcome, the latter capturing players' beliefs about how much more or less than the equitable payoff the player is allocating the other players (the player's 'kindness function') and about the reverse relationship (how much the other players are being kind to her). It can be noted that here again, in order to be able to derive a tractable solution,

the outcome is assumed to be determined according to the calculation made by the players to maximize their utility.

Fehr and Schmidt (1999) propose instead to model fairness as "self-centered inequity aversion" by integrating relative payoffs into the utility function. Contrary to Rabin's model, they do not take into account players beliefs about each other but only the difference in monetary payoffs. The utility function thus defined is a combination of the absolute value of the player's monetary payoffs and a fairness component which weights the relative payoff difference to the other player. Two independent parameters weight the payoff difference according to the sign of this difference, allowing for a distinction between the two opposite sources of inequity aversion, advantageous and disadvantageous inequity. Importantly, a player's utility is assumed to possibly be negatively affected by a difference to their own material payoffs only, which differentiates inequity aversion as modeled here from altruism, the latter implying that subjects do care about others' situation regardless of their own.

Starting from the same premise that a player's utility also depends on their relative standing, Bolton and Ockenfels (2000) propose a model that can be considered as a generalization of the previous one. Considering  $n$ -player games, they define the inequity aversion component according to the difference between the player's monetary payoff and the *average* monetary payoff to the other players. Their formalization however implies that advantageous and disadvantageous inequity aversions are not differentiated, which experimental results tend to contradict, and that players take into account only the average payoff to the group, which is equivalent in two-player settings but may be hard to justify with more players since in that case players would be indifferent to highly uneven outcomes as compared to egalitarian outcomes, as long as their payoff fits with the average.

Falk and Fischbacher (2006) present a combined model that builds on both intentionality (as Rabin's) and on outcomes to mitigate the monetary payoff. Similarly to Rabin (1993), they include in the utility function a 'kindness term', which is itself composed of an 'outcome term' capturing inequality aversion and weighted by an 'intention factor' grasping the degree of intentionality of the outcome, as well as a 'reciprocation term' determining the extent to which the player modifies the payoff to the other player(s) according to their previous perceived fairness.

All these models extend the behavioral assumptions used in the game theoretical resolution of non-cooperative games, allowing for motives other than monetary payoff maximization. This can increase the accuracy of predictions in bilateral interaction situations and leaves

room for cooperative outcomes that do not rely on incentive mechanisms. However, originating from game theoretical analysis, they fall short of integrating into subjects behavior the broader considerations that can be raised by collective production processes. Moreover, some of the experimental results that they seek to explain, and that they may indeed claim to predict to a greater extent than narrow self-interest models, may also find satisfying alternative explanations relying either on the mere effect of repetition on agents holding broader beliefs about others' behavior, or on the actualization of other behavioral drivers: social norms, habits, intrinsic motivation. The latter might also provide a broader frame of analysis of collective production processes in which most of the elements from the diverse approaches presented above could be integrated. This would however constitute a tremendous task that cannot be achieved through this thesis, and would exceed by far the scope of experimental economics. We will thus rather briefly present them to set the background of our narrower endeavor.

## **7. Looking forward: Norms, institutions, habits and motivation**

Discussing the influence of other-regarding preferences and motivation on collective production processes is less of an attempt to bridge contested terrain than a challenge to tidy up an over-crowded field, where diverse very rich traditions meet, mingle, sometimes merge but seldom let a clear picture emerge. Since a comprehensive synthesis would largely exceed the scope of the present chapter, we will only briefly present here a few theoretical approaches that could form an integrative basis for the study of individual behavior in collective production processes. Rather than postulating a complete and deterministic model of choice, including both choice processes and motivations (or *ends* in Simon's words) and assuming an optimizing program, it seems more reasonable to favor hypotheses that can be generated by the observation of regularities in human behavior and which leave open the possibility of heterogeneous behavioral patterns upon which the diverging theories presented above build.

### **7.1. *Social norms and social preferences***

Though concerns of reciprocity and equity are generally discussed under the label "social preferences", especially within the field of experimental game theory, social norms at large have been considered since long<sup>24</sup> as behavioral drivers arising from repeated interactions

---

<sup>24</sup> See e.g. Summer (1906).

among human agents and more generally as necessary regulators of human societies<sup>25</sup>. Collective production processes, as embedded in broader social interaction processes, can thus be assumed to be shaped or, more loosely, influenced by the behavioral norms existing when and where they take place.

As defined in Granovetter (1985) social embeddedness is "the extent to which economic action is linked to or depends on action or institutions that are non-economic in content, goals or processes". As an illustration of the importance of embeddedness in the description and understanding of economic processes, Granovetter (2005) quotes the deleterious effects of corruption, but also the "less often noted, but probably more important, savings achieved when actors pursue economic goals through non-economic institutions and practices to whose costs they made little or no contribution. For example, employers who recruit through social networks need not - and probably could not - pay to create the trust and obligations that motivate friends and relatives to help one another find employment. Such trust and obligations arise from the way a society's institutions pattern kin and friendship ties, and any economic efficiency gains resulting from them are a byproduct, typically unintended, of actions and patterns enacted by individuals with noneconomic motivations." Though a strict classification into economic and noneconomic activities may be hard to justify and to implement, this stance on social embeddedness can be used to show that factors that have been removed by some analytical traditions from the study of economic processes may well play a major role, shaping agents motivations and explaining choices that cannot be understood from the sole consideration of the so-called narrow 'economic', or monetary interests.

Granovetter's words on the implication of group size on the effectiveness of norms are strikingly reminiscent of a decisive part of Olson's analysis, when he notes in the same paper that "collective action that depends on overcoming free-rider problems is more likely in groups whose social network is dense and cohesive, since actors in such networks typically internalize norms that discourage free riding and emphasize trust. Note that all else equal, larger groups will have lower network density because people have cognitive, emotional, spatial and temporal limits on how many social ties they can sustain. Thus, the larger the group, the lower its ability to crystallize and enforce norms, including those against free riding. The insight that free-rider behavior is especially unlikely within immediate families is a special case of this argument" (Granovetter, 2005). As in Olson's work, the norm is

---

<sup>25</sup> Durckheim (1973).

described here as a modifier of the self-interested valuation, adding non monetary rewards to the balance.

However, since social norms tend to have a spontaneous, unwritten and informal character<sup>26</sup>, their influence on individual behavior might not be limited to the mechanisms of sanctions and rewards stemming from social embeddedness. Norms can also be enacted *for their own sake*, or maybe more precisely as a component of more or less fundamental levels of the representation of the world. Cook and Hardin (2001) discuss the distinction between exclusionary and universalistic norms. Universalistic norms are not associated with sanctions in the form of exclusion from a particular group, and among the universalistic norms they further distinguish norms covering "dyadic or small group interactions" and those covering "essentially collective interactions", the latter being "less likely to be backed by interests independent of sanctions than are universalistic dyadic norms". Once a norm is internalized by individuals the reward associated to compliant behavior is generated autonomously<sup>27</sup>. If the norm coincides with cooperative behavior the ensuing cooperation must be considered as emerging from this spontaneous choice, and the corresponding collective production is achieved through the mediation of an internalized norm.

## 7.2. *Institutions and habits*

It could be argued that norms, especially internalized ones, are special cases of institutions. In Veblen's words, "institutions are, in substance, prevalent habits of thought with respect to particular relations and particular functions of the individual and of the community" (Veblen, 2003, p.127).

Habit, in turn, is a propensity, a "response disposition that is activated automatically by the context cues that co-occurred with response during past performance" (Neal et al., 2006, p.198). Presented by Aristotle<sup>28</sup> as forming the "second nature" of man, shown by Ravaissou (1838) to be the necessary link between "nature and consciousness", habits have long been ignored as idealism and rationalistic views of human behavior were playing down any form of restriction on free deliberation. Similarly, *homo economicus*, or in Meckling's (1976) wording "REMM" - for "resourceful, economizing, maximizing man" - seems to be exerting its willpower in a social vacuum and out of an empty personal history. However, as suggested in Hodgson (2004b), the role of habit, as "submerged repertoires of potential

<sup>26</sup> Hechter and Opp (2001).

<sup>27</sup> Horne (2001).

<sup>28</sup> In Aristotle (1953), also suggested in Aristotle (1992) as the source of virtue.

behaviour" determining or driving actual behavior "does not make belief, reason or will any less important or real". Following the early institutionalists (see Hodgson, 1998) and grounding human behavior on habit instead of rational choice acknowledges the strong shaping effect of social interactions and personal history on the various determinants of individual behavior including, crucially, reason, belief and the "rules of the game" (Hodgson, 2004b, p.653 and Hodgson, 1998, p.182), without denying the importance of deliberation and purposive choice. As summed up by Pascal (1963) "we are as much automatic as intellectual ", hence the "two extremes: to exclude reason, to admit reason only ".

In a dynamic perspective, "habits emerge from the gradual learning of associations between responses and the features of performance context that have historically co-varied with them" (Wood and Neal 2007, p.843). Given the heterogeneity of performance context, through history and throughout the various cultures, societies and civilizations that human beings have built and which have shaped individual behavior, habits appear to be far from deterministic structures, while at the same time they seem to constitute the source of strong behavioral drivers that allow human agents to economize on cognitive resources. This crucial role of habit as a determinant of individual behavior further entails and underlines the large diversity of potential behavior and the decisive influence of both context and personal history on individual choices. Moreover, habit and hence behavior are shaped by institutions so that observed behavior can only be interpreted as a clue about the situation as perceived by the subject and not as evidence of universal behavioral rules. Accordingly, individual decisions occurring in collective production situations can hardly be accurately predicted by unitary models that assume the highest level of homogeneity among agents and implicitly consider contextual and psychological factors as irrelevant<sup>29</sup>. Of course, these models may still capture a part of the picture, which can in turn easily be checked through experimental procedures.

### ***7.3. Intrinsic motivation vs. extrinsic incentives?***

On the other side, knowledge-based theories of the firm have sourced their behavioral assumptions on considerations of social embeddedness, communities and social preferences, and more generally insisted that incentive mechanisms were not an adequate tool to manage organizational learning processes. Nevertheless, monetary incentives, peer pressure, reputation effects and community participation as a way to enhance locally available knowledge can all be considered to some extent as external sources of motivation.

---

<sup>29</sup> This is also the conclusion in Gibbons (1998) : "One simple possibility is that economic models that ignore social psychology are incomplete (but perhaps useful) descriptions of incentives in organizations."



Reciprocity motives, which have been argued to be an important component of individual choices, involve a bilateral exchange where the focus is not on the content of the activity itself but on its perceived benefit and thus also imply an external or indirect source of motivation. Concerns of fairness and equity, viewed as the enactment of social norms<sup>30</sup>, also draw on external motivators linked to the action rather than on the behavior itself.

However, the extensive developments of the knowledge-based theories of the firm, drawing on organizational studies, led researchers in economics and management science to revisit the behavioral assumptions of contract theories which were not able to adequately address the properties of knowledge creation settings and organizational learning processes. The cognitive dimension of collective production processes, implying knowledge creation and problem-solving activities, can indeed not be reduced to the simple trade-offs occurring in the case of information exchanges. Meanwhile, motivation studies had been given a renewed attention by psychologists (Deci and Ryan, 2000, p.256). Though discussions and integration efforts have focused on the so-called 'crowding-out' effect of rewards on intrinsic motivation, which was also the starting point of this stream of literature in the psychological field (Deci, 1971), the subsequent efforts to analyze the determinants of motivation, the various types thereof and their effect on task performance led to the integrated self-determination theory (SDT).

Formalized in Deci and Ryan (1980, 1985, 1987 and 2000), first under the label CET (for 'cognitive evaluation theory'), SDT originates from numerous field and experimental observations. SDT owes its name to the crucial role of autonomy, as opposed to control, on determining the type of motivation at stake and thus the effect of the institutional arrangements thereon. Rewards, but also threats, deadlines, evaluation and surveillance are typical examples of controlling contextual factors, events that will possibly trigger the undermining effect on intrinsic motivation. According to the perceived locus of causality (PLOC), on a continuum from external to internal, various types of motivation can be defined. Whereas intrinsic motivation is necessarily related to intrinsic regulation and an internal PLOC and may thus unravel if controlling factors like incentives are introduced, extrinsic motivation can be associated with self-determined, autonomous behavior, or on the contrary to controlled behavior depending, among others, on the degree to which the action is instrumental to other self-determined behavior.

---

<sup>30</sup> Other-regarding preferences may also be considered as expressions of the need of relatedness emphasized by the theories of intrinsic motivation .

SDT postulates that the three psychological needs for relatedness, competency and autonomy determine the level of intrinsic motivation. Accordingly, it does not suggest that task may be ranked along the motivation type continuum, but rather that contextual and environmental factors do play a crucial role in the motivation at stake and thus in the agents' performance.

Considering the determinants of intrinsic motivation it seems obvious that a large part of productive activities do not fall into the category supporting its presence. However, the various types of extrinsic motivation supporting an internal PLOC and which can thus be internalized allow to integrate a whole range of actions and activities which agents may not find interesting *per se* but which they will nonetheless pursue for the sake of the general attributions pertaining to their position.

The SDT does not only underline the role and importance of intrinsic motivation but also sets a full continuum from amotivation to intrinsic motivation through extrinsic motivation, on which activities can tentatively be placed but also which, maybe more interestingly, supports a finer perception of the interplay between internal and external sources of motivation. Far from precluding the existence of extrinsic motivation and the role of incentive as an extrinsic motivator, it offers an integrative framework for analyzing both controlled and autonomous behavior and their various determinants.

## **8. Conclusion: irreconcilable views?**

Contract theories and organizational learning theories provide a strikingly different picture of collective production processes. On one side, individual agents doted with independent production capacities set up incentive schemes that allow them to coordinate when information asymmetries would lead market transactions to fail. The prediction of failure is derived from the short-sighted and monetarily determined choice process. On the other side, individual agents are chiefly considered as elementary components of the organizational processes that 'encapsulate' the collective knowledge of the organization, in which they engage on a long term basis.

These two contrasting views can however be argued to be complementary to the extent that they address the same processes. Certainly, their contradictory behavioral assumptions cannot be reconciled as given since they constitute one-sided, exclusive representations situated at the two opposite edges of the motivational continuum. Whereas organizational learning theories relegate incentives to a second order, black-boxed question, which at best

has to be solved on the fly while focusing on the knowledge creation processes that sustain the core activities of the organization, they constitute the very core of the conclusions stemming from the behavioral hypotheses of the contract theories. It is thus no wonder that these two streams have either ignored or criticized each other and that to date no integrative discussions have been engaged and pursued. A broader reconsideration of the questions at stake may nonetheless give some clues to the benevolent and persevering one about the possibility of fitting these unfitting pieces of the puzzle together. Let's consider that the contract theories are built on the five following assumptions:

- 1) agents choose the alternative that maximize their own payoffs,
- 2) agents use or choose as if their were using unlimited computational resources,
- 3) agents are solely interested by and solely take into consideration monetary payoffs,
- 4) agents are myopic and only consider immediate outcomes,
- 5) agents believe that all other agents have the above decision rules and preferences.

Letting aside the two first hypotheses, which have already extensively been discussed in the literature and concern the decision process rather than the behavioral motives, we can argue that the three remaining ones constitute the core source of incompatibility between the two alternative approaches. We can observe that extending the types of returns considered by agents (assumption 3), including consideration of forward periods (assumption 4) and assuming that agents may hold a larger range of beliefs on each others than the ones defined in the two previous assumptions (assumption 5) already greatly extend the behavioral framework to leave room for cooperation in many more settings, opening the way for the sustainable contribution by organization members required by knowledge creation processes.

The experimental observations reviewed in the next chapter as well as the original contributions presented in the second part of this thesis specifically focus on assumptions 3 and 4, and provide some clues on assumption 5.

---

## CHAPTER II

### ILLUSTRATING THE DOUBLE EDGE OF COOPERATION

#### LESSONS FROM EXPERIMENTAL STUDIES OF SOCIAL DILEMMA

---

##### Summary

We review experimental results from various social dilemma settings. Diverse mechanisms have been used to observe cooperative behavior in experimental environments and generated robust results. Our first intention is to present a review of the main results yielded by various game settings addressing the influence of interaction structures and parameter values on individuals' decisions and on group outcomes in terms of cooperative behavior. Far from claiming exhaustiveness we rather wish to draw the main conclusions about the determinants of cooperative outcomes from a very rich literature sometimes originating from quite different motivations. The two alternative conceptions of cooperation - a behavior to be elicited from diverging interests and an emergent property stemming from social interactions among agents - find support from the laboratory experiments, in a large range of settings involving two or more players interacting in a symmetric or in an asymmetric frame.

## 1. Introduction

In this chapter we will review experimental results from various social dilemma settings. Diverse mechanisms have been used to observe cooperative behavior in experimental settings and have so far generated a series of stylized facts which can be considered strong results. Our first objective is to present a review of the main results yielded by various game settings, exploring how these various designs and parameter values affected players' choices and group outcomes and thus weighting their relative influence on cooperative behavior. Because of the tremendous amount of experimental work generated by the study of social dilemma situations we limited the scope of this review to the major findings, which have proven to be robust throughout many conditions.

Social dilemma experiments cover a range of experimental settings which, though described under different labels, share the common trait of offering subjects a choice between (i) cooperating at the risk of being outperformed by others but for higher potential collective payoffs and (ii) not cooperating (or "defecting") securing their own payoff at the cost of a lower potential collective performance. The variety of the designs used allows testing for as many various determinants of cooperative outcomes. Whether two or more players are involved, whether their interactions are symmetric or not and whether they can choose between alternative mechanisms or are constrained in one type of frame of interaction, the experimental designs reviewed here offer a subtle and contrasting picture of individual behavior. As it will be discussed later *no single behavioral hypothesis implying homogeneous agents can explain the diversity of the observed results*, and one may consider it as hopeless to try and infer behavioral models for prediction purposes. However, some conclusive regularity can be observed concerning both the individual behavioral drivers and the comparative performance of alternative settings. Taken together these results tend to support the heterogeneity of individual choices and its persistence over time. Since many of the on-going debates on cooperation and on design choices reflect diverging views on individual decision-making it is suggested that cooperative outcomes depend on some fit between stakeholders and the proposed institutional mechanism, and can thus not be derived *in abstracto*.

*The two visions of cooperation - a behavior to be elicited from diverging interests and an emergent property stemming from social interactions among agents - find support from the laboratory experiments.*

Economic experiments have some common guidelines, standards of the art, which must be kept in mind when elaborating on the data analysis. One of these imperatives is to provide an incentive framework crucially determined by the monetary type of rewards that subjects may get. Whatever choices are made, the outcome always boils down to a difference in monetary payoff. Accordingly, results analysis solely relies on data generated and controlled within the incentive framework.

Reflecting theoretical advances, time and repetition have been introduced as alternative treatment in almost all kinds of social dilemma experiments. Time is not an easy factor to reproduce and repetition alone cannot account for time length and for history. Experiments necessarily take place within a limited time span since one cannot ask subjects to stay in the room for more than one or two hours without considerably reducing the interest of subjects in participating, or their availability. 'Partner' treatments, where each player is always matched with the same other player(s), have been introduced to isolate the mere effect of the repetition of the game that occurs in 'stranger' treatments where players are matched with different players in each repetition. More generally, experiments necessarily exclude from the parameter space, by definition, those that can not be reproduced within the experimental framework because they are crucially dependent on time (history) or because they cannot be introduced in controllable experimental settings.

So what 'cooperation' are we talking about? Experiments allow to test either immediate factors favoring spontaneous cooperative choices, or hypotheses about recurrent behavioral patterns that support the emergence of cooperation at the group level, when variations on monetary pay-offs are at stake. Though experimental results demand to be taken to the field, and confronted to full-fledged observations, they can offer a fine and dynamic picture of the sensitivity of the collective outcome to a large set of conditions as well as some hints about the heterogeneous underlying individual motivations.

The remainder of the chapter is divided along the two main dimensions of the general structure of the games, namely (i) the relationship between players situations (symmetry vs. asymmetry) and (ii) the number of players (two vs. more than two). Following the degree of complexity involved we present symmetric (sections 2 and 3) before asymmetric (section 4 and 5) game situations, and in each case we move from two-player to multi-player settings. Any division of this kind is obviously bound to draw some arbitrary lines between closely related settings, in which cases we have tried to mention and restore the disconnected link. Section 6 offers a tentative summary of the various behavioral drivers highlighted on the way.

## 2. Symmetric two-person interactions: prisoner's dilemma experiments

### 2.1. *Stylized facts: cooperation happens*

The structure of the prisoner's dilemma game was first used by Melvin Dresher and Merrill Flood (Flood, 1952) as a way to evaluate the robustness of the very concept of equilibrium that Nash had just developed, a couple of months before Albert Tucker framed the game as a situation where "two men, charged with a joint violation of law, are held separately by the police" (Tucker, 1950). In what seems to be the very first prisoner's dilemma experiment, Armen Alchian and John Williams performed one hundred repetitions of the game. The cooperative outcome was reached in 60 occurrences, the Nash equilibrium in 14 occurrences, and Alchian did trick Williams 18 times against 8 the other way round. However, both players there were somewhat familiar with zero-sum game theory and the paper mentions "an accident at the outset" that prevented anonymity to be fully implemented.

In a more controlled environment, Rapoport and Chamah (1965) used seven alternative parameter values for the payoffs (i.e. seven different games), and ran three partner treatments varying the number of games played by each pair and the order in which the different games are played. They crossed this setting with a "no matrix" condition where players were given information only about the relative payoffs ("win" or "lose") and not the value of the payoffs. The overall results show that the cooperative choice ("C") was observed in 46% of all cases, all games and all treatments taken together. They observed a striking difference between the two information conditions, with a much higher frequency when players are shown the whole payoff matrix than when they know only about the general payoffs structure and the actual payoff values (60% vs. 30%). Even in games where the payoff matrix involved a one-to-fifty rate between potential loss and possible gain frequency was 35% and 21% respectively with and without payoff information. In the subsequent chapters, Rapoport and Chamah draw a fine picture of the dynamics observed in the repeated game. First, they noticed a general, strong tendency of players to imitate their partner's choice, pointing that "the product moment correlation of the frequencies of cooperative responses of paired subjects is in many cases very nearly plus one" (p.66). The other important feature explaining the interaction dynamics is "a tendency not to persist in the rewarded defect ("D") response (a cooperative bias) and a tendency to persist in the punished D response (a noncooperative bias)" (p.86). In other words, the repeated prisoner's dilemma game provided a first illustration of the role of reciprocity in cooperative choices. The combination of the strong correlation between paired players choices and the cooperative bias effectively led to

"the increasing predominance of CC and DD states" over unilateral states, and to the general dynamics of cooperation choices, characterized by an initial trend towards more defection followed by a "recovery" when "the frequency of cooperative responses increases" (p.102) up to its initial level.

Running forty rounds of a prisoner's dilemma game under three conditions with low, high and asymmetric payoffs, Beckenkamp et al. (2007) found remarkably similar results. In the asymmetric condition, payoffs from the low and the high payoffs condition were crossed and applied to either of the two players. Over all rounds the cooperative choice was made in 70% of cases in the low payoffs condition, 59% in the high payoffs condition and 39% in the asymmetric condition. In all three conditions, cooperation rates furthermore followed a similar pattern, starting around 60% in the first round, declining down to 42% in the second round, recovering after round eight and stabilizing around 80%, 65% and 55% respectively. They also found mutual cooperation to be more stable than the other states, and mutual cooperation choices to be more frequently followed by further mutual cooperation choices in symmetric than in asymmetric conditions.

## 2.2. *The sequential prisoner's dilemma game*

Shafir and Tversky (1992) report results from a series of games where players are asked to make their choice in three different conditions: a condition where they know that the other player has chosen to cooperate, a condition where they know that the other player has chosen not to cooperate (has chosen to 'compete') and a third condition where the game is played simultaneously. They observed 37% of 'cooperate' choices in the simultaneous condition, 16% when players are informed that the other player has chosen to cooperate and 3% when they are informed that the other player has chosen not to cooperate. Shafir and Tverski interpreted the paradoxical behavior of players who chose to cooperate in the simultaneous condition but chose not to cooperate in either case of the sequential condition as "quasi-magical thinking", a belief that their choice may influence the other player's choice when it has not been done at the time they make their own choice. They also suggested that it may be partly attributed to "a shift from individual to collective rationality" because the group optimum is more salient in the simultaneous setting. The latter explanation could somehow be connected to the hiatus between 'espoused' theory and 'theory-in-use' as presented by Argyris and Schön as the basis of their theory of action<sup>31</sup>. The argument put forward by Shafir and Tverski is indeed that "due to the presence of uncertainty [people] may not see

<sup>31</sup> A complete exposition of their model can be found in Argyris and Schön (1978).



their own preferences very clearly", or in other words that they choose according to a general norm of what they think to be the appropriate behavior - espoused theory - and not according to the same driver that determines their choice under certainty – theory-in-use. Though their settings used only one payoff matrix, this 'non consequential', incoherent result must be kept in mind when cooperation choices are analyzed in one-shot, simultaneous-play settings.

Clark and Sefton (2001) implemented sequential prisoner's dilemma games in a stranger setting where subjects played ten repetitions. They also found that first movers, who played in a similar situation to the simultaneous game, chose the cooperative option more often than the second movers (42% against 23% of all choices in round one and 28% against 9% in round ten). The cooperative option was also chosen more often by second movers when first movers had chosen to cooperate, amounting to 41% of choices in the first round and 23% in the last round. Finally, doubling the payoff associated with a defective choice when the other player had chosen to cooperate ('Double Temptation' matrix) or doubling all stakes significantly reduced cooperation rates.

In a more complex setting, Potters et al. (2007) varied the information available to the second mover of a two-player game which amounts to a prisoner's dilemma game for one of the three possible parameter values. In that condition, and when the parameter value was common knowledge, first movers chose to cooperate in 16% of cases in the simultaneous treatment, and in 27% of cases in the sequential treatment, where second movers were also informed of the first movers' choice (in the simultaneous treatment second movers would still choose after the first movers, but without being informed about their choices). In accordance with the previous findings presented above, second movers cooperative moves drop from 27% down to 9% when they are informed about the first movers' decision before making their own choice.

### ***2.3. Opening the prison: voluntary participation***

On the ground that "outside the prison, humans usually don't have to interact with each other", Orbell et al. (1984), Orbell and Dawes (1993) and Hauk (2003) added an exit option in their respective settings. Keser and Montmarquette (2003) also let the two players chose between a two-person prisoner's dilemma game and a "private remuneration" where only a miscalculation could lead to a non-optimal pay-off. In most of these settings, the highest rates of cooperative choices were observed when players could choose whether to participate in the game or not. These results are further reviewed next chapter (Section 2), where we present an experiment applying voluntary participation to a public good game setting.

Though the prisoner's dilemma game is the most reduced version of social dilemma, involving only two players and a binary choice, the significant share of cooperative choices observed in the numerous replications of the experiment and with most parameter sets strongly suggests that the emergence of cooperation cannot be ruled out and that the factors favoring this spontaneous behavior need to be carefully studied. We thus now turn to the extended versions of social dilemma provided by the n-prisoner's dilemma experiments, where players still face a binary choice, and by the larger class of public-good game experiments where players typically choose their contribution level on a quasi-continuous scale.

### **3. Symmetric, n-person interactions: public good game and n-prisoner's dilemma experiments**

Public good game experiments were first designed by social psychologists and sociologists in order to explore the determinants of the provision of voluntary provision of public goods and identify what helps groups overcome social dilemma. Work from economists was rather motivated by the continuing dispute on problems arising from misrepresentation of preferences when estimating the demand for public goods (Bohm, 1972). The extensive and unavoidable review of J. Ledyard, published as a chapter of the Handbook of Experimental Economics (1995), describes in details the experimental procedures used during the first two decades of this effort and offers a clear synthesis of consensual results that have emerged along the way. We will thus only briefly comment on this part of the literature, in order to pinpoint specific factors particularly relevant to illustrate the emergence of cooperation and address the advances made in the identification of the determinants of cooperative outcomes since the time of the publication of the Handbook.

Though public good game settings are very similar to prisoner's dilemma games and likewise feature several agents interacting without differentiated roles, they provide a richer frame of interaction involving more than two players. They also allow to broaden the range of cooperation levels among which players can choose and thus to generate fine-grained observations on individual and parameter effects.

### 3.1. *A misleading stylized fact: positive, decaying contributions*

Though many experimental results from public good game experiments are still waiting to be gathered and tabulated one can confidently expect a first round or one-shot average contribution of around 50% of the initial endowment, followed by a significant decay over time in the repeated setting<sup>32</sup>. The voluntary contribution mechanism (VCM) thus displays some efficacy and the figure is all the more significant than choices resulting from all kinds of player strategies are included. Would free-riding be a universal dominant behavioral pattern one should observe an average contribution level well below one half of the initial endowment. It seems thus to be the case that the 'first move' choice of the experimental subjects involved is quite variable and that behavioral heterogeneity is high enough not to be overlooked and not to be considered only as a marginal, negligible phenomenon.

The first strand of explanation for these puzzling, theoretically inconsistent<sup>33</sup> observations had to do with theories of errors and learning. Here learning is simply about agents acquiring some knowledge about the dominant strategy equilibrium. In other words, repeating the game should help agents identify the equilibrium after a few initial mistakes. This could be done in two possible ways: either it takes a few rounds before the player really 'steps into' the game setting and gets the point or the other players' choices force him back to the dominant strategy choice - teach him the right strategy. Some cooperation may thus be observed in one-shot games but it would fade away after a few repetitions while beliefs and expectations are updated with experience.

The usual picture displaying the evolution of the average contribution level, averaged over all groups, apparently supports this explanation. A more detailed analysis of experimental results through a closer look at the group level average contributions suggests however that slightly different observations can be made. The data from the experiment presented in Chapter 3 of this thesis, for instance, clearly show that contribution levels dramatically vary from one group to the other. In the Baseline treatment, featuring a standard, linear public good game, two of the nine groups exhibited no significant decreasing trend (Page's trend test, p-value respectively 0.7 and 0.5), for an average contribution level of 25% and 44% over all twenty rounds. Similarly contrasting with the picture offered by the aggregate results, contribution rates above 50% were observed in four groups in rounds as late as the last two or three ones. Though the social optimum of full contribution has been reached in

---

<sup>32</sup> To date, and to our knowledge, no public good game experiment featuring the VCM in its original form has yielded significantly different first round outcome on average contribution levels.

<sup>33</sup> Inconsistent with the free-riding hypothesis.

none of these cases, it is clear from these observations that contribution levels observed in first rounds can be sustained throughout all rounds of repeated settings. As summarized by Hichri and Kirman (2007), "the decline in aggregate contributions cannot be explained by resorting to a uniform model of individual behavior".

Adding support to this interpretation of the aggregate trend, Andreoni (1988) and Croson (1996) both observed that an unexpected restart of the game after an initial sequence triggered a sharp increase of contribution levels in the first rounds of the following sequence.

The remainder of this section presents several factors which have been observed to influence the rate of cooperative choices or the contribution level, from characteristics of the game environment to refinements of the interaction mechanism itself, and summarizes the corresponding results to draw a precise picture of the conditions under which spontaneous cooperation can be most expected to emerge.

### **3.2. *Parameters: Group size and multiplier value***

Within the typical settings of a public good game experiment two parameters can substantially modify the stakes (and thus the features of the experimented situation): the number of players and the 'productivity' of the collective project or 'marginal per capita return' (MPCR). Since experiments have historically been run with much fewer players than involved in the real situations under study the effect of group size has also been of major concern. The tricky point here is that one cannot vary the first without changing the second, and more generally it is quite a challenging exercise to vary one parameter of the game keeping other things equal.

Group size *per se* can generate contradictory effects on contribution levels. Olson (1971) makes the point that "the greater effectiveness of relatively small groups [...] is evident from observation and experience as well as from theory". Considering meetings that involve "too many persons" (i.e. too many to reach a fast and effective decision) he notes that "though all of those participating presumably have an interest in reaching sound decisions, this all too often fails to happen". The reason is that "when the number of participants is large, the typical participant will know that if own efforts will probably not make much difference to the outcome, and that he will be affected by the meeting's decision in much the same way no matter how much or how little effort he puts into studying the issues" (p.53).

In strict terms of beliefs and expectations about other players' behavior, however, larger groups could be expected to deliver higher contribution rates. If the return on investment of the collective production function (MPCR) is kept constant, the average contribution of the other members of the group necessary for one player to get the same payoff decreases with the size of the group. The private incentive to contribute is still negative when group size is increased but positive contribution decisions needs lower expectations on the contributions of the other members of the group.

As summarized in Holt and Laury (2008), contributions have experimentally been observed to increase with group size for low (0.3) MPCR, to substantially increase with MPCR for small group sizes ( $N=4, 10$ ) and to vary little when either group size or MPCR is kept large ( $N=100, MPCR=0.75$ ). The group size effect is however dominated by the MPCR effect when the product of group size and MPCR is held constant.

### 3.3. *Communication*

Among all the alternative treatments that have been tested, treatments allowing for basic interpersonal social interaction have proven to be the most effective in generating sustained, high levels of cooperation. Dawes et al. (1977) were among the first to test the effect of communication on cooperation in multiple-player settings. Running  $n$ -prisoner's dilemma games where only if all players chose to contribute the cooperative choice led to a higher monetary payoff, they found that communication increased the share of cooperative choices, from 30% in the no-communication setting up to 72% when communication was unrestricted. In the 'irrelevant communication' setting where players were not allowed to discuss the experimental situation, an average 35% only of cooperative choices were observed, pointing to the role of communication in creating a shared perception of the interaction situation at stake rather than a mere socializing tool. The results from Isaac and Walker (1988b) are quite as clear-cut. In a follow-up study expanding Isaac and Walker (1988a) they investigate the impact of face-to-face communication as a "means of altering free-riding behavior in the context of the voluntary contribution mechanism". When the four subjects of each group were given four minutes to discuss qualitative aspects of the experiment, contribution levels were found to be dramatically higher than in the reference treatments without communication. Moreover, no significant decrease was observed in the following periods. No less striking is the outcome of the C/NC treatment, where communication was implemented in the first ten rounds, and suppressed for the ten remaining ones (these are results from their Design I where endowments were symmetric in

that all subjects get a equal endowment). Average contributions, expressed as a percentage of the optimum level - when all subjects invest all of their endowments - reached 100% after a few rounds and stayed this high until the very last rounds. Even the end-effect does not level down contributions below 80% of the group optimum. This observation tends to attract attention to another aspect of group size, which is the indirect effect it can have on cooperation level through communication. Whatever explanation is favored for the difference in contribution levels achieved by groups of various sizes such a dramatic effect of communication implies that interaction structures play a major role in the determination of cooperative outcomes. Cooperation should be most observed in settings where communication is facilitated and these are intuitively thought of as implying reasonable group size, all other factors being equal. Hence the question of achievable levels of cooperation may not be adequately tackled if some prior considerations about the characteristics of the group under consideration are not included in the analysis. We will come later to the arguments about agents' heterogeneity but we can already make a point on group characteristics stemming from the mere number of agents belonging to the group at stake. In the context of public good experiments groups typically involve four or five players and large group settings have usually been implemented with ten players (as in Isaac and Walker, 1988a). Isaac et al. (1994) used the largest groups to our knowledge, comparing treatments with groups of 40 or 100 subjects. Obviously larger figures would already be too high to implement sustained, face-to-face communication and in real settings groups of this size are integrated organizations relying on local cooperation within differentiated subsets<sup>34</sup>.

Orbell et al. (1988) tried to go further in the identification of the exact interplay between communication and induced cooperative outcomes. Several media can indeed be invoked in the explanation of the positive effect of discussion between the involved actors on contribution choices, and their experiment is a first attempt to distinguish more precisely between (i) promotion of generalized cooperative norms, (ii) commitment through promises and (iii) in-group bias induced by the minimal group paradigm<sup>35</sup>. They could reject the first factor, based on their observation that discussion fostered cooperation levels mostly towards the players of the groups within which communication had happened. Contribution rates

---

<sup>34</sup> Here again we are reminded of Olson's remark on 'confederations' or large groups composed of differentiated smaller subgroups. His argument about the possibility for small groups to achieve collective goals is also strengthened by the results about the effect of communication reported in this section, though with a different tone since the experimental settings involve only a minimal breach of anonymity that cannot compare with the 'social incentives' described by Olson.

<sup>35</sup> On the minimal-group paradigm, see e.g. Tajfel and Turner (1979)

however rose from 37.5% up to 78.6% when communication was introduced, in groups where subjects had to deal only with their own group, supporting a strong effect of pre-play communication on cooperation levels. Furthermore, in all settings they observed higher cooperation rates when contributions were directed to the group to which the subjects were initially assigned as compared to when contributions were directed to players of the other group, confirming the intra-group bias created by the mere belonging to a distinct group.

Ostrom et al. (1992) explores the effect of communication under the experimental design labeled 'covenant without a sword'. Running a common pool resource (CPR) game they observed also higher contribution rates when communication was possible, sustained contribution rates when communication was repeated, and noted that lower values of endowments led to higher contribution rates.

Frohlich and Oppenheimer (1998) focused on the effect of communication mode and compared the results of face-to-face and e-mail communication. Subjects in the communication treatments contributed on average 87% of their endowment, against 29% in the no-communication treatment. Their face-to-face treatment reproduced the results from Isaac and Walker (1988b) with a contribution level of nearly 100% over all rounds, whereas the e-mail treatment was significantly less effective, though it yielded an averaged cooperation rate of 76% of the social optimum.

On the same streak of effort to disentangle the various effects of communication on voluntary cooperation rates, Bochet et al. (2006) ran three treatments of a public good game that involved three different modes of communication, namely 'face to face' (FF), 'chat room' (CR) and 'numerical cheap talk' (NCT). Whereas the baseline (B) treatment generates a somehow standard average contribution rate of 63% in the first round, 47.5% over all ten rounds and a consistent decay, 100% contribution rates have been observed in most rounds of the FF treatment, where subjects could talk for 5 minutes prior to the first round, and 96% averaged over the 10 rounds. In CR, where discussion was mediated through an on-line chat room, subjects being free to post messages censored only for "threats and offers of side-payments, revealing one's identity, and obscene language", an average 81.5% contribution rate was observed, with interestingly no round exhibiting full contribution. The average contribution in NCT, where only contribution levels could be typed by the subjects during the discussion phases, was 47% over all rounds. Though both FF and CR treatments yielded significantly higher contribution than NCT, pointing again to the crucial importance of the content of discussion and of the depth of the non-binding interaction, average contribution levels were still 18% higher in FF than in CR (though significant at the 10% level only). In

other words, both communication mode and content are important factors influencing cooperation rates.

### 3.4. *Commitment, promises or 'covenants without a sword'*

As suggested and explored by Orbell et al. (1988), the effectiveness of communication in enhancing contribution levels could be attributed to the medium of commitment and promises that such settings allow the subjects to make. Their results suggest that promises to cooperate, made during the discussion, can be binding to the extent that all players in the same group promise, and that some group identity can be created during the discussion period. Contrarily to Hobbes<sup>36</sup> and to the vision stemming from game theory analysis that any non-binding discussion is cheap-talk<sup>37</sup>, they could thus conclude to the possibility that "covenants *are* parts of cooperation's explanation and that, therefore they are *not* but words."

Ostrom Walker Gardner (1992) opens with the same quotation of Hobbes'. They also observed that promises of cooperation were made and extracted by subjects when communication was possible. Though their setting implied an interior social optimum and thus made the task of determining the optimal contribution level more difficult to achieve for the experimental subjects, all groups managed to reach an agreement during the discussion round in the one-shot communication treatment. The agreement proved to be effective since average payoffs rose to 74% of the highest possible amount yielded by the social optimum, from a highest rate of 32% in the final rounds of the baseline treatment. When communication was repeatedly made available throughout the whole game the figure reached 99% in the setting with the lower endowment, as compared to 35% without communication, while it reached about the same value as in the one-shot communication treatment (73%) in the higher endowment setting. The communication rounds were effectively used not only to correctly determine the optimal contribution level, to reach agreements and make and extract promises to abide by these agreements, but also to pressure unknown defectors into fulfilling them whenever defection was detected.

However, commitment may not be the main factor explaining the effect of communication on cooperation levels. The detailed results in Dawes et al. (1977) interestingly show that though cooperative choices raise to 84% when unrestricted communication was followed by a non-binding vote, this holds only in the condition where (substantial) loss is possible.

---

<sup>36</sup> "Covenants, without the sword, are but words, and of no strength to secure a man at all" (Hobbes 1909, p.85, as quoted in Orbell et al. 1988).

<sup>37</sup> See e.g. Farrell and Rabin (1996).



Otherwise the cooperation rate decreases to 58%, i.e. to a lower value than in the no-vote unrestricted communication condition. Players choosing not to cooperate after reaching some level of certainty about other players cooperative choices should not be a complete surprise any more after the results from Shafir and Tversky (1992) and Clark and Sefton (2001) on sequential prisoner's dilemma games (see section 2.2 above).

### 3.5. *Endowment value and symmetry*

Isaac and Walker (1988b) also presented the results of a condition where endowments are asymmetric. Within the groups of four subjects, two subjects were endowed with 54 tokens while the two others received 70 tokens. Though the difference between the symmetric and the asymmetric conditions is only marginal during the 10 rounds where communication is allowed, contribution rates fall sharply in the final 10 rounds without communication when endowments are asymmetric, down to less than 15% from the third round on.

Van Dijk and Wilke (1994) ran a one-shot public good game where half the subjects in groups of four were endowed with 30 points and the other two subjects with 10 points. They observed that the average contribution of the latter was lower (6.6 points against 16.4) though their contributions expressed as shares of the endowment were on average higher (66% against 55%). In a treatment where the difference in endowment was presented to the subjects as stemming from a difference in time investment, the average contribution of the low-endowment subjects reached 7.4 (74%) whereas the the high-endowment subjects contributed an average of 14.1 (47%). Since the average contribution rate over all conditions was 60%, which is within the range of contribution rates observed elsewhere in VCM experiments, and since they did not run a control treatment with symmetric endowments, it is not clear that endowment asymmetry had any effect on cooperation levels. The most interesting result here is the effect of the perception by the subjects of the origin of endowment asymmetry, namely that when the difference in endowments was associated with a difference in merit (time invested), the average amounts contributed by the two types of players were closer.

Still in the context of a one-shot linear public good game, Cherry, Kroll and Shogren (2005) found that subjects in groups with heterogeneous endowments contributed less (33% of their endowment) than subjects in groups with homogeneous endowments (42%), the social optimum being also here for each subjects to contribute all of their endowment. In the heterogeneous endowments, each of the four subjects in one group was attributed a different endowment. Comparing contributions along two conditions where subjects in both the

asymmetric and the symmetric endowment treatments earned their endowment either according to their performance in a GMAT test or through random allocation, they found no significant difference in average contributions between the two conditions across all subjects. However, contrary to Van Dijk and Wilke (1994), subjects with lower endowments in heterogeneous endowment groups contributed more when the endowment was random than when the endowment was stemming from a difference in merit, assessed through a task performance. Similarly, subjects with higher endowments in the same groups contributed more when the endowment was earned according to performance. These seemingly contradictory results point to the crucial importance of context-specific cues for the activation of cooperative behavior. It appears here, for instance, that relating uneven endowment distributions to time spent by subjects, which is most likely not attributed to subjects' personal competence, and relating them to task performance, which most likely is, work opposite.

Buckley and Croson (2006) designed a repeated public good game where two of the four subjects of a group got an endowment of 25, the two others getting an endowment of 50. They also provided subjects, after each round, with the average accumulated wealth of the other subjects in their group. They could thus compare the relative effect of wealth and income discrepancy. Their first result confirms the deleterious effect of endowment asymmetry on average contribution levels, which were significantly lower than in similar experimental settings with symmetric endowments. They also observed that low-endowment players contributed on average the same amount to the public good, in absolute value, as high-endowment players, which obviously translates into higher contributions rates, expressed in percentage of the endowment, from the low-endowment players. The same relationship was observed between "wealth" as measured by cumulated earnings and contributions.

Among the above experiments, in the ones featuring repeated settings (Isaac and Walker, 1988b, and Buckley and Croson, 2006) endowments were kept constant throughout all the repetitions of the game. In Maurice et al. (2011) intra-group measures of the effect of asymmetric endowments were possible through a change in individual endowments within the game, in the setting of public good game with an interior Nash equilibrium. In their "equal distribution" or symmetric condition all four players were endowed with 20 tokens whereas in the "unequal distribution" or asymmetric condition two players received 15 tokens as endowments and the other two 25 tokens. After 10 rounds under either the asymmetric (I) or the symmetric (E) condition, a second series of 10 rounds was announced

where conditions were switched - which the authors questionably labeled "income redistribution" – or maintained, leading to the four treatments EE, II, EI and IE. They found that groups with asymmetric endowments in the second series contributed the same average amount as groups with symmetric endowments, whether the first series had been run under the symmetric condition or not. Between-subjects analysis showed that subjects with lower endowments contributed relatively more than subjects with higher endowments, the former over-contributing to the Nash equilibrium contribution level whereas the latter Nash-contributed, which confirmed the results of the previous experiments.

Similar results were also found in Keser et al. (2011). When endowments were unevenly distributed amongst the four players of a group but only weakly so, players endowments being respectively 10, 15, 15 and 20, no significant difference in average contribution levels was found with the symmetric benchmark setting where all players got 15. When endowments were distributed on a one-to-four ratio, three players getting 8 and the fourth payer getting 36, contribution rates were substantially lower all along the 25 rounds. Here again, more precisely in the condition where asymmetry was strong, lower-endowment players contributed a higher share of their endowment than higher-endowment players, both contributing on average the same absolute amount.

The effect of asymmetric endowments thus somehow contrasts with the results on reciprocating behavior and on inequality aversion in public good game experiments, which can be suggested by the strong pattern of adjustment towards the observed contribution of the other players. In asymmetric settings lower-endowment players indeed tend to contribute on average a higher share of their endowments than higher-endowment players, and doing so exhibit no clear aversion to inequity. Voluntary contribution levels have thus been observed to depend crucially on the identification by the actors of the very situation at stake, as cued by the broader interaction context.

### ***3.6. Provision points and thresholds***

Provision point and threshold experiments relax the assumption that all players must contribute their whole endowment if the social optimum is to be reached. Since it has been observed that stakes, and in particular the absolute and relative differences between the free-riding and the social optimum contribution, can have a substantial effect on contribution levels, non-linear public good games may trigger diverging results.

*Critical mass*

In one of the earliest experimental attempts to study social dilemma, Marwell and Ames (1979) used a setting in which the individual payoffs depended on a step function of the total contributions by the group members, including a provision point where the return on investment (MPCR) of the public good sharply increases. Since they observed an average contribution of 57% of the total endowment, well above the free riding level, they controlled for a possible threshold effect in their follow-up study (Marwell and Ames, 1980). They report a 51% average contribution level in the latter setting without MPCR jump, finding no significant difference between the two conditions and thus no significant effect of the provision point.

*Provision point experiments*

Proper provision point experiments, as reviewed in Croson and Marks (2000) involve the provision of a public good only if the sum or the number of contributions is higher or equal to a predetermined threshold. Van de Kragt et al. (1983) developed a threshold game in which subjects can choose to make a fixed contribution to the group account or not. If at least, respectively, 3 and 5 players in a group of 7 chose to contribute, the public good was provided. In the condition where no communication was allowed, the public good was provided in respectively 72% and 61% of the groups. Dawes et al. (1986) investigated the effect of providing the players with a money back guarantee in case the threshold is not met. Depending on the higher or lower cost function of the public good, they observed that either 40% or 70% of the groups were successfully financed the public good when no money back guarantee was available. In the condition involving a money back guarantee, the minimum contribution rate was achieved in respectively 57% and 100% of cases. In a similar setting where the game was repeated, Rapoport and Eshed-Levy (1989) observed that the public good was provided in 34% of all rounds in the benchmark treatment without money back, and in 54% of all cases when a money-back guarantee was provided in that contributing players got the same payoff as non-contributing players, tantamount to the endowment, if the public good was not provided. These treatments with money-back guarantee offer an interesting observation on the role of risk-aversion on contribution choices. They show that risk-aversion is indeed a significant factor explaining some of the non-contribution choices, though it can only partly explain the failure to achieve social optimum. Rapoport and Eshed-Levy (1989) implemented a third treatment without money-back guarantee but where, if achieved, the public good yielded the same payoff to both contributors and non-contributors.

The success rate rose to 72% of all cases, indicating a stronger effect of pay-off maximizing behavior compared to risk-aversion in determining contribution levels. These "binary contribution" threshold experiments allow to draw a clear picture of contribution decisions in one-shot settings, when players cannot fine-tune their contribution level. Since players then face fewer alternatives they cannot use partial contribution as a trade-off between the risk of low payoffs because of others' free-riding choices and the will to contribute so as to be in coherence with the socially optimal choice. Moreover, successful production of the public good is achieved to the risk of over-contribution if the number of contributing players is higher than the minimum required.

Threshold experiments have also been conducted in the more common setting of continuous contributions, where participants can allocate each unit of their endowment to either a public or a private account. Isaac et al. (1989) used an repeated partner experimental design based on different provision points ("Low", "Medium" or "High") and also tested for the introduction of a money-back guarantee. The first pertinent observation from their results can be made on the High provision point treatment, where all subjects of a group had to contribute all their endowment for the public good to be produced. Only one group out of six achieved the social optimum, first in the second round, but then achieved it again for all the eight remaining rounds. Thus, if the contribution level averaged over all groups followed the usual pattern, starting around 60% and rapidly decreasing to 25%, the introduction of a threshold was sufficient, in one case, to reach and sustain the social optimum. Contrastingly, all the groups in the Medium and Low provision points treatments (total contribution of respectively 87% and 44% of total players endowment required for the public good to be produced) failed to reach the social optimum, in all rounds. All three treatments displayed the usual average contribution levels, which mechanically implies that the number of groups where (at least some suboptimal level of) the public good was produced increased in lower provision points, in the earlier rounds. In the condition where a money-back guarantee was implemented both High and Medium treatments displayed an averaged 86% contribution level, sustained throughout all ten rounds. Moreover, 67% of the groups of the High treatment achieved the social optimum in the five last rounds, and 57% of all group choices were socially optimal.

### 3.7. *Sequential contributions and leadership*

Erev and Rapoport (1990) compared simultaneous and sequential choices in a binary-contribution provision point public goods game. Three out of five players had to choose to contribute in order for the public good to be provided. In line with previous findings on the effect of risk aversion on contribution levels, they found that the average contribution level is higher if the public good game is played sequentially. In their simultaneous treatment, the public good was produced in only one out of seven groups, 43% of all players having chosen to contribute, whereas ten out of the fifteen groups in the sequential condition managed to produce it, with 75% of subjects choosing to cooperate. Figuières et al. (2012) observed a similar outcome in a continuous-contribution experiment without threshold. In the sequential treatment, where players chose their contribution level in a randomly determined order, and where each subsequent player is informed of the amount contributed by previous players in their group, the average contribution reached 52% against 41% in the simultaneous condition. Their design was set in order to differentiate between the mere effect of sequentiality as such and that of the information that higher ranked players get about previous choices of their group members. The positive effect of sequentiality was observed only when subsequent players were informed of previous contributions. Importantly, the observed positive effect could be attributed to the first movers, last movers contributions being lower than the average contribution in the simultaneous setting. Norman and Rau (2011) also report results from a repeated, public-good game but with only two players. They run both simultaneous and sequential treatments with either one or two provision points and also observe higher rates of public good provision in the sequential condition, with 80% of groups achieving the first provision point against 62% in the simultaneous choice treatment.

In Güth et al. (2007) only one player is selected each round to play before the other members of the group make their choice. Combining two conditions where the leader is either fixed for all rounds or rotating from one round to the other, they observe higher contribution levels (54% of the social optimum over all rounds) than in the benchmark treatment where all subjects contribute simultaneously (40%). Though they observe a strong correlation between leaders' and followers' contribution rates, their results also indicate that followers only partially react to the initial contribution of the leaders and contribute significantly less than the leaders do. This gives further evidence of the actual but only partial role of risk-aversion and uncertainty about other players' choices in non-cooperative behaviors. Moreover, contribution levels significantly decrease over time in both treatments.

### 3.8. *Refinements of the mechanism 1 : sanctions and rewards*

Since it has been argued that in real settings agents not only interact on a repeated basis but are also in a position to reciprocate on cooperative or non-cooperative behavior, the VCM has been refined in numerous experiments in order to allow for rewards and punishments of players by other players, based on the observed contribution levels. This is a major variation from the standard public good setting where the absence of such a mechanism is justified by the non observability of preferences rather than by non observability of choices. In the case of collective production processes, however, preferences for the achievement of the common product are assumed to be obvious and participants can indeed observe to a reasonable degree their respective contribution levels.

Introducing the possibility for group members to sanction other members after contribution information is circulated within the group, Ostrom et al. (1992) observed only a partial effect on cooperation rates compared to their baseline treatment where no sanctions are available to players. Efficiency of the group choices however increased from 21% up to 37% of the social optimum. When the possibility to impose a fine on other players is combined with communication, however, it stacked with the positive effect of communication with an average cooperation success of 85% of the optimal level.

Running partner and stranger treatments of a standard public good game, Fehr and Gächter (2000) crossed two conditions where subjects were either given or not the opportunity to punish each other after being informed of the individual contributions of the other group members. In the stranger setting, where group composition changed randomly from one round to the other, they observed sustained average contribution rates of 58% of the social optimum during the ten rounds when punishment was available, against 18.5% without punishment. These rates rose respectively to 85% and 37.5% in the partner treatment, where 82.5% of all subjects contributed their whole endowment. In both treatments, contribution rates exhibited the usual decaying trend in the no-punishment condition whereas it either stayed constant or even increased over time in the punishment condition. In a similar experiment, Masclet et al. (2003) also observed that the availability of costly, monetary punishment positively influenced contribution rates, from an average 27% of subjects endowments up to 55%. Moreover, they found that the non-monetary sanctions condition, where costless "disapproval points" could be attributed to other players without affecting their payoffs, also triggered higher contributions though to a lesser extent (45% of the social optimum). Their setting added a final series of no-punishment rounds after the initial series

crossing the no-punishment and punishment conditions. In all treatments, the final series exhibited very similar contribution rates to the first no-punishment condition, ruling out any sustained effect of the punishment rounds.

In Ertan et al. (2005) the choice to introduce or not a punishment device was left to subjects through a vote. Though the introduction of a vote mechanism made their settings somewhat more difficult to interpret, their results also clearly show higher contribution levels in groups that voted for the possibility of sanctions. In that case, average contribution was above 90% and increasing over time. The number of players voting for the introduction of a punishment device also increased over time. In groups where players chose not to make punishment available, the usual decrease of contribution levels was observed, excepted when the vote was repeated five times over the whole time span, starting before the first round.

Sefton et al. (2007) tested separately for sanctions and rewards as well as for a combination of both. Compared to their baseline treatment where players contributed on average 44% of the social optimum in the last but one ten rounds<sup>38</sup>, the introduction of costly sanctions yielded an average 60% contribution rate. In the reward treatment the average rose to 64%, and up to 75% when both rewards and sanctions were available. Not surprisingly the efficiency, measured by group earnings as a percentage of the amount generated by the social optimum, is higher in the reward treatment since rewards are in the form of zero-sum transfers between the rewarding and the rewarded player. In the sanctions and rewards treatment, efficiency is equally high, rewards being used more often than sanctions. In all three rewards/sanctions treatments, the total amount used as rewards and/or sanctions decreases over time although contribution rates are quite stable. This stability in spite of the decreasing use of endogenous incentive mechanisms may point to the settlement of a common understanding, indicating that the mere availability of the device is a coordination tool that allows cooperative players to implement cooperation standards.

Bochet et al. (2006) crossed their four treatments with a condition where players had the opportunity to punish other group members. Adding costly sanctions to their baseline treatment merely affected the stability of contribution levels over time. In the baseline treatment average contribution started at 63% of total endowments and declined to 20% in the last round for an average of 47.5% over all rounds whereas it started at 70% when sanctions were available and stayed about this level in all but the last rounds where it fell to

---

<sup>38</sup> In order to isolate the end effect we computed the average contribution rates over rounds 11 to 19 only. In the ten first rounds all groups played the standard public good game without sanctioning or rewarding opportunities.



61%. Among the three communication treatments, given the high contribution levels triggered in the 'face to face' (96%) and 'chat room' (81%) ones, only the 'numerical cheap talk' treatment was significantly affected, in a very similar way to the baseline since contribution rates also rose from 47% up to 67% over all rounds, the latter being sustained in all but the last rounds. In the CR treatment, the average contribution level was slightly increased, from 81% up to 96% when sanctions were available. Given the cost of sanctioning, however, earnings were extremely close within each pair of conditions (with and without punishment), which points even further to the crucial role of communication. *Punishment, sanctions and rewards within groups bound to collective action can thus mostly be considered as default devices aimed at sustaining some minimum amount of cooperation but deprived of decisive enhancing effect.*

#### *A different type of sanctions: exclusion*

In the experiments presented above, sanctions and rewards were in the form of monetary transfers or 'approval points' that never affected group composition. In real settings, the exclusion from a group, an organization or even a territory has often been considered as the utmost form of sanctioning.

Cinyabuguma et al. (2005) offered subjects playing a public good game in groups of sixteen the opportunity to vote to remove players from the original group, after each round and after being shown other players' individual current and past contributions. The expulsion vote was costly if the outcome was an effective exclusion following the majority rule, and the expelled players would be moved to a secondary group for the remaining rounds. In all settings where the expulsion vote was available, contribution rates were above 90% in most rounds, in the group of non-expelled players. Contrastingly, contributions in the expelled players' group rapidly declined to zero. On average, only three players were expelled in each group, but the effect of effective expulsion stacked with the effect of the threat of expulsion, since players who had been 'nominated' for expulsion without receiving enough votes did increase their contributions in following rounds. Altogether, contribution rates and efficiency as well were greatly increased by the possibility of expulsion.

The settings of Güth et al. (2007) also included treatments named 'strong leaders' for the first contributing player, the leader, can exclude one of the three other players for the next round after observing their individual contributions. The average contribution, which rose from 40% in the benchmark treatment to 54% when a leader without exclusion power was chosen, reached 80% in the conditions where exclusion was possible. The addition of the sanction

mechanism in the form of exclusion power also affected the sustainability of contribution levels, which are stable and around 80% throughout the sixteen rounds. As in Cinyabuguma et al. (2005), the threat of expulsion from the collective process was effective in raising contribution levels of non-excluded players.

These result about the relative efficiency of the availability of sanctions in the form of temporary exclusion or definitive expulsion mechanisms however rely on the possibility to apply them after each round, which may not correspond to the usual situation of collective production processes, and did not completely control for the cost of reducing group size when players are excluded.

### ***3.9. Refinements of the mechanism 2 : manipulating matching patterns***

#### *Exogenous group formation*

In all the experimental settings described above, subjects were randomly assigned to the group within which they interact. This allows to make straight treatment comparisons and test the aggregate relative outcome of the diverse settings explored. The observed cooperation rates are thus robust to subject heterogeneity and general indications can be derived from the corresponding results. However, matching patterns in human societies are only partly random and often rely on some mutual knowledge shared between agents concerning their behavioral drivers. It may thus be argued that matching players according to their display of similar patterns of choice instead of randomly, or even letting them choose their partners, could help replicate a crucial aspect of the interaction environment.

Following the classification of player types used in Fischbacher et al. (2001), and also using the strategy method as part of their procedure, Burlando and Guala (2005) formed groups of players according to the result of a series of classification tests. After the classification process, players tagged as either free-riders, cooperators or 'reciprocators' were matched with players of the same tag. In the homogeneous groups treatment, groups formed of cooperators contributed on average 82% of their endowment, groups formed of 'reciprocators' 93%, while groups formed of free-riders only 11%. Contribution levels in cooperators groups exhibited a moderate though significant decline over time, whereas 'reciprocators' groups kept their contribution rate higher than 95% in the first sixteen rounds (out of twenty). Not surprisingly given the identification process of players types, free-riders contributed slightly less in the homogeneous than in the heterogeneous condition where groups were randomly formed among all players, cooperators contributed only slightly more, whereas 'reciprocators'

sharply increased their contributions in all rounds. In a similar flavor of exogenous matching process, Ones and Putterman (2007) created groups according to the contribution choices observed in a preliminary round and to monetary punishment choices of the four following rounds. The result was somehow stable contribution rates over the ten rounds when groups had been randomly formed, which is in line with the effect of monetary punishment presented in the previous section, and they also observed that groups formed with players displaying both highest contribution levels and highest tendency to punish free-riders achieved the highest average contribution rates, with 94% of their total endowment against 58% for groups formed of the lowest ranked players.

### *Endogenous group formation*

A few other public good experiments have implemented endogenous group composition. Page et al. (2005) ran 20 rounds of a 2x2 design, crossing monetary punishment and endogenous regrouping. In the regrouping condition, subjects had the opportunity after each third period to rank the fifteen other subjects participating in the same session according to their preference to be matched in the same group in the following rounds. The ranking choice was costly, and informed by the contribution history of the other players. The four lower ranked (most preferred) subjects were then matched in the same group and the computation was repeated twice until all subjects were rematched. Contribution rates were higher when either punishment or preferential rematching was available (about 70% in both cases) than in the baseline treatment (38%), and highest in the combined treatment when both were implemented (about 80%). In the latter case, the average contribution level was furthermore sustained over most rounds. In each of the four first groups of the regrouping treatment, composed of the four 'most preferred' subjects of each session, the overall average contribution topped 90%, with more than 80% of the subjects in these groups having remained in the same group until the last round.

Other attempts to implement endogenous group formation are somehow trickier to analyze. Both Ehrhart and Keser (1999) and Ahn et al. (2008) allowed subject to change groups from one round to the other. Both settings allowed for group size to vary from one to the whole "population" (respectively 9 and 12 subjects). In Ehrhart and Keser (1999) switching group was costly for players but no other restriction was applied to their choice. Subjects first played a three-person public good game, and had the opportunity at the beginning of every other round to chose either to stay in his group, to join another group or to go in a new group by himself. Players' decisions were informed by all past contributions and size of all groups.

They observed a continuous flight of higher contributing players away from lower ones, the latter then chasing them in turn in their new group, triggering decreased contribution levels and again the flight of higher contributors. All in all, contribution levels displayed the usual declining tendency, indicating that the free-riding pattern is strong enough to deter high contributions if groups cannot isolate from the whole population. In Ahn et al. (2008) subjects started in groups of one, and costless rematching requests were either uncontrolled or submitted to a vote of the players remaining in their previous group. The vote concerned either the entry (restricted entry/free exit treatment) or the exit (free entry, restricted exit treatment) decision. When entry was restricted, conditioned by the result of the group members vote, repeated denial of entry led to increased contribution levels.

### ***3.10. Group incentives and team competition***

Addressing "the rapid growth of group incentive compensation systems in the United States between 1945 and 1991", Nalbantian and Schotter (1997) report experimental results comparing various group incentive schemes. Running two consecutive conditions of twenty-five rounds each they crossed a 'revenue sharing' scheme - a form of voluntary contribution mechanism with an interior Nash equilibrium and where the collective output is partially determined by a random variable - with (i) 'forcing contracts', (ii) 'profit sharing', and (iii) competitive teams' schemes. They ran separately an individual 'monitoring' scheme, crossing a high and a low probability for monitoring to happen. In the 'forcing contract' scheme the total revenue of the group is shared among the members only if it meets an exogenously set outcome target, otherwise each member gets the same fixed 'penalty wage'. This condition is very similar to the provision point settings presented in section 3.6 above, the penalty wage being the equivalent of the players endowment. The 'profit sharing' scheme sets a lower target to the group but provides no wage if the target is not met. In the 'competitive scheme' pairs of groups are formed and a transfer is made to the group that achieves the highest output. Finally, in the monitoring condition subjects payoffs depend only on their own contribution level and on a probability to be controlled. In all condition, the effort cost function makes losses possible. Though detailed results are only partially presented in their paper, they tend to indicate that revenue sharing, profit sharing and forcing contracts schemes elicited on average the same level of cooperation, the two former ones starting on average around 35% and the latter around 45% of the social optimum, decreasing over time. The dummy variable regressions coefficients show that though the forcing contract treatment displays a positive and significant effect on effort levels for the five first rounds, it has a

strong negative effect on the average cooperation level in the final five rounds and thus does not perform better than the two others over the whole time span. Given that the effort level of the Nash equilibrium are higher in the forcing contract (75) and in the profit sharing (19.1) schemes than in the revenue sharing condition (12.5), it seems that the latter outperformed the two other ones in terms of group net surplus, and that the first one failed to elicit the targeted cooperation level. Monitoring was highly efficient but only when a high probability (0.7) is set for monitoring to happen. In the low monitoring probability (0.3) condition effort levels were the lowest among all treatments. Finally, competitive teams achieved around 40% cooperation levels, sustained over time.

Tan and Bolle (2007) ran a more standard public good game, repeated over ten rounds. They introduced intergroup competition, with and without incentives. In their 'XPG' setting, each group member was informed at the end of the round whether their group total contribution was lower or higher than that of another group. In the 'CPG' treatment a higher (lower) multiplier is used in the payoff function of the highest (lowest) contributing group. In all three treatments the pure strategy Nash equilibrium is to free-ride. In both partner and stranger conditions the average contribution rate was higher in the CPG and XPG treatments than in the baseline treatment without group comparison. Moreover, group comparison with incentives elicited stable cooperation levels over time in the (more usual) partner treatment, at 40% of the highest possible level.

All in all, the numerous public-good game designs presented in this Section provide evidence that the observation of positive contribution levels proves robust even among anonymous players and can be enhanced by empirically consistent factors. We will now turn to games featuring asymmetric structures, where players face different initial situations.

#### **4. Asymmetric two-person interactions: principal-agent settings, the 'gift-exchange game' and the ultimatum/dictator games**

Experimental principal-agent settings involve an asymmetric role for players. In both two-player and multiple-player cases one of the subject in each group is in the position of a principal whereas the others are agents. They make different kind of decisions in a sequence of proposition and response. The context is typically one of the "gift exchange game" type as in Fehr, Kirchsteiger and Riedl (1993), using effort level for the agent instead of quality level. The principal picks up a contract and proposes it to the agent(s). The contract is in the

form of a payoff function depending on the agent's effort level or a fixed wage, associated in some cases with a fine to be paid if the agent is found to choose an effort level lower than a specified desired minimum level. In the standard design the agents then reply by accepting or not the contract and if so by choosing an effort level.

#### **4.1. *Incentives, control and trust***

The first experiments of this kind were designed as tests for the role of reciprocity in economic interactions, as in Fehr et al. (1997) or in Fehr and Gächter (1998). The treatments thus did not include incentive schemes but contract offers that specified a wage and a desired effort level. "Workers", upon acceptance of the contract, choose the actual effort which could differ from the expressed desired one. The cost function for the agent's effort is set such that effort cost is increasing with effort level. The contract is thus incomplete in the sense that the wage is paid by the employer (or "firm") to the worker whatever effort level the latter chooses. In this setting agents have been found to reciprocate, in typically providing higher effort levels when offered higher wages. This result is at odd with the game theoretical solution assuming selfish, profit maximizing players which predicts that workers will choose the minimum effort level and employers, anticipating this best response of workers, will propose the minimum positive wage. Experimental results clearly indicate that employers anticipate that agents will provide a higher effort level and usually ask for a desired effort of 7 (on a scale from 1 to 10), together with an offered wage that almost split the total income generated if the agent actually chooses this effort level. It appears that this anticipation is far from being undecieved by the workers. Most of them do not fully meet the desired effort level but choose levels substantially above the selfish choice of 1 (with an average of 4.4 out of the same scale from 1 to 10). Thus cooperation arises also in this setting of gift exchange games, without communication and without the implementation of costly binding contracts. These results already raise serious concerns about the effect of incentive schemes since they may interact in a non neutral way with a predisposition to cooperate. More generally it emphasizes the potential for the emergence of cooperation at least in settings involving two-person interactions. It is even more interesting since these outcomes have been observed in one-shot or in 'stranger'<sup>39</sup> experimental settings where no anticipation (resulting on diverging beliefs about the future behavior of other players or on signaling behavior) can be assumed excepted the short-term anticipations of the principals (employers or buyers) concerning

---

<sup>39</sup> Reminder: in a 'stranger' treatment the game is repeated but players are randomly matched each round, as opposed to a 'partner' treatment where players are matched once and for all rounds before the first round.

agents' (workers or sellers) acknowledgment of the 'appeal to cooperation' sent by the principals.

The usual serious concerns can be raised about the generalization of the experimental results, which may imply among others a parameter jump on stake levels. One of the interesting points in this strand of experiments is that the regularities observed appear to be robust against stake level variations (Fehr and Tougareva, 1996). The occurrence of outcomes where appeals to cooperation are sent by the principals and positively responded by the agents can thus be stated as cases where anonymous agents, meeting only once in a context that offers a clear non cooperative pay-off dominant choice for a high-stake amount of money, manage to reach cooperative outcomes without the implementation of (costly) coercive mechanism.

In a follow-up study, Falk and al. (1999) find principals' offers to start at around 45% of the maximum possible wage and agents to reply with an effort level chosen on average at 50% of the maximum available (around 45% in the stranger treatment and 55% in the partner treatment). Contrarily to what happens for instance, in public good game settings, effort levels in the partner treatment even increased over time. Over all ten rounds average efforts level reached 73% of the maximum available and more than 80% in the final round.

Here again we can observe strikingly divergent behavioral schemes within the same settings, with some agents clearly not positively responding at all to any appeal to cooperate and sticking to the dominant strategy of minimum effort level whereas others exhibit persistent cooperative choices, positively responding to 'offers for cooperation' that can even elicit higher levels of cooperation if the game is repeated between the same principal/agent pairs.

### *Introducing incentives*

Fehr and Gächter (2002) investigate the effect of the introduction of incentives in the basic design. They report the results of an experiment designed to test the net effect of incentives contracts compared to the 'trust' contracts described above. In the Incentive Treatment (hereunder IT) the contract offer made by the principal includes a fine to be paid in case of verified shirking. If shirking occurs, in that the worker chooses an effort level lower than the desired one specified in the accepted contract, the worker has to pay the amount of the fine. A random draw determines with a probability of one third whether the effort is verified. The question then in this setting is whether the cooperative outcome previously observed in the Trust Treatment (hereafter TT) will be enhanced by the introduction of an incentive scheme

or if the frame of interaction proposed in the IT triggers different choice patterns. In other words, will the call for cooperation be better heard if it is associated with the threat of a potential punishment? Or can the mere existence of a coercive tool affect the potential for voluntary cooperation?

As we have seen, incentive systems are typically considered as a way to enforce incomplete contracts. When control on the agent's action is available, it has been shown theoretically (see e.g. Holström, 1979, Grossman and Hart, 1983, Hart and Holström, 1987) as well as experimentally (see e.g. Prendergast, 1999 and Chiappori and Salanié, 2003 for reviews) that properly designed incentive systems can generally be said to have some positive effect on the principal's profit. This is a quite standard result. However, organization theorists know well about the impossibility to rely solely on contracts to get things done. Even in the context of a stranger treatment, in the TT they implemented Fehr and Gächter found again effort levels to be substantially higher than the minimum effort requested by the principals. The average effort levels were furthermore significantly higher in the Trust Treatment than in the Incentive Treatment. Moreover the effort levels chosen by the agents in the IT were never higher than the effort levels specified by the principals, in incentive compatible contracts (i.e. in IT contracts where the expected value of the fine is higher than the cost of the desired effort), contrasting with the positive correlation observed in TT between offered rents (offered wage minus desired effort level) and chosen effort levels. In TT around 68% of agents chose effort levels higher than the best response prediction. This merely means that voluntary cooperation completely vanishes in IT as compared to TT. Interestingly, average fines are almost maximal in the IT, with 69% of all IT contracts specifying maximal fine levels. This heavy use of the incentive tool goes along with less reliance on high wage offers and with lower desired effort levels. Average wage is half higher in TT than in IT and approximately the same difference is observed in median desired efforts. The net outcome for principals is that their surpluses (income from agent's effort level minus wage paid to the agent) are higher on average in IT. It seems thus that it is more expensive to 'appeal to cooperation' than to use the credible threat of a fine though the total surplus generated is higher in TT.

The implementation of incentives in the form of a fine can be pointed out as potentially creating a negative frame, which would considerably lessen the accuracy of any comparison between the two treatments. In order to estimate the strength of this potential framing effect Fehr and Gächter ran a third treatment, the Bonus Treatment (BT). In this treatment incentives are implemented in the form of a bonus not to be paid to the agent if she is caught



to be shirking. The probability of verifiable shirking is, as in IT, one third. Thus the BT can be described as a positive reframing of the IT and any substantial difference in results between TT and BT denotes a substantial difference between TT and IT net of framing effect. It appears that voluntary cooperation is much higher in BT than in IT with a positive correlation between offered rents and chosen effort levels observed in BT, suggesting the presence of a strong framing effect. Still, the amount of extra effort provided by agents is significantly lower in BT than in TT, supporting that the introduction of an incentive mechanism was enough to entail a substantial break down of cooperative behavior.

#### 4.2. *Contract choice*

Up to this point, the gift exchange game experiments from the time of the ones that have been reported so far make no alternative institution or mechanism available to the players. It can however be argued that in real settings not only do players interact within a given institution providing them a frame for interaction but they also do often find themselves in situations where they must or can make a preliminary choice determining the mechanism or institution that will support their transaction. This can be the case either when concerned agents are picking up a contractual form or when agents are looking for counterparts to match with according to the kind of contract they propose, or even by choosing a social environment more akin to their preferences.

In the context of the gift exchange game Fehr, Klein and Schmidt (2007) report the results of an experiment where they allow players to choose between alternative contract types. They ran stranger treatments including ten repetitions of the stage game. In a first treatment (the Trust-Incentive treatment, hereunder TI) the principal makes a prior choice between the two kinds of contracts implemented in Fehr and Gächter (2002) treatments TT and IT: the Trust Contract (TC) and the Incentive Contract (IC). Practically they could specify a fine to be paid in case of verified shirking, or only a wage and a desired effort level without enforcement device. Recalling that in the previously described TT the average profit of the principal was lower than in the IT it is no surprise that in a situation where the two kinds of contracts compete, incentive contracts are observed to largely dominate trust contracts and that the share of incentive contracts increases substantially over time. Furthermore, when chosen incentive contracts significantly outperformed trust contracts in terms of average effort levels and principal's payoff. Despite these observations however, the number of trust contracts never completely declined down to zero and accounted for between 20% and 30% of all contracts from the fourth round on. A tentative conclusion here could be that the

positive correlation between offered wages and voluntary cooperation from the agents is not strong enough to be widely relied upon by principals. Since many principals who offered trust contracts in the first rounds had their cooperative offers not fully reciprocated by the agents they rapidly switched to incentive contracts. One important aspect of real setting interactions however is the repetition of similar transactions between the same agents. In Fehr et al. (2007) all treatments were implemented under stranger condition, in which players are matched with different partners from round to round so that they never play twice in the same dyad. It should also be noted that in all the experimental principal-agent settings mentioned in this sub-section the principals are always the proposers, the responding agents having thus no voice in the design of the contract and can only accept or refuse the proposed contract.

The contrast between the observations reported from previous experiments under TT and the poor performance of trust contracts against incentive contracts is however striking. It looks like principals and agents failed to coordinate on a cooperative outcome, the former not proposing highest wages for fear of being matched with a non-cooperative player and the latter responding with effort levels low enough to make the incentive contract more attractive. One explanation may be that in this setting principals were bearing all the 'cost of trust' since they had no mean to signal their willingness to cooperate before making their final decision. The second treatment implemented in Fehr et al. (2007) somehow balances the situation in giving the final word to the principals: in this Bonus-Incentive treatment (BI) players choose between the same Incentive Contract as before and a Bonus Contract (BC). The BC implemented here is different from the one available in the Bonus Treatment in Fehr and Gächter (2002). The BC here is a generalization of the Trust Contract used in the Trust-Incentive treatment, since principals can offer to pay a bonus after agents have chosen their effort levels, without being bound to pay it no matter which effort level is actually chosen. If they choose a BC they must specify a wage, a desired effort level and a bonus they intend to pay if the agent chooses an effort higher or equal to the desired level. Principals can choose a null bonus, in which case the chosen contract is similar to the Trust Contract used in TI. Over all rounds, the BC was chosen in 85% of cases and a bonus was offered and paid in 80% of these cases when the agent chose an non-minimal effort level. It appeared that some principals did experiment the IC for a few rounds after choosing the BC in the two first rounds and then switched back to BC. After round five (out of ten rounds) the BC was always chosen in more than 85% of cases, and the percentage of BC went up to 96% in the last round.

All in all what lessons can we draw from this strand of experiments? First, the standard predictions relying on selfish and myopic payoff maximizing players are seldom observed excepted in the choices of effort levels in the (negative) incentive treatment. Second, even in one-shot-like stranger treatments substantial levels of voluntary cooperation are observed. The most chosen contract, which also outperformed any others in terms of individual as well as collective surpluses did not include any coercive mechanism but was based on the optional payment of a fixed bonus. Third, when alternative contracts can be chosen, none of the available alternatives happens to eventually be driven out by the other one and especially trust contracts relying solely on voluntary cooperation steadily represented one fourth of all chosen contracts after the initial decrease of their share.

However, the experimental settings presented in this section all involve two-person interactions, and it is a widely observed result that size *per se* can have a substantial effect on behavioral regularities.

#### **4.3. *Monitoring intensity: extrinsic incentives and intrinsic motivation***

The designs of the principal-agent experiments reviewed above implied that the adjustment variables on the principals' side were the wage and the desired effort level, the intensity of monitoring being fixed. Though it was most often observed that agents' average effort levels were not lower in the diverse 'incentiveless' conditions, this outcome could partly be attributed to principals' choices on wage. In Dickinson and Villeval (2008) settings, contrastingly, where agents had to perform the basic task of progressively uncovering the curve of an increasing function during a "one-minute 'work' period", wages and required effort levels were exogenously chosen by the experimenters. The adjustment variable for the principals was the intensity of monitoring. Agents' choices departing from predictions based on myopic selfish solutions could thus mostly be attributed either to a combination of their willingness to achieve some equity in the relative payoffs and their willingness to achieve a higher level of completion of the task (in the 'variable' treatment) or solely to the latter, tantamount to intrinsic motivation (in the 'fixed' treatment). The variable treatment indeed made the principal's payoff always increasing with the agent's effort whereas in the fixed treatment only the achievement of one or both of the two (exogenously set) effort thresholds was relevant in the determination of the principal's payoff. They observed that around one third of the agents in all treatments set their effort at a higher level than the one required to achieve the maximum possible payoff, some of them at their own cost, confirming at once

some degree of spontaneous cooperation, some reciprocity as well as the general interest of the subjects in performing the task and thus some intrinsic motivation. Although their results show that monitoring intensity had a positive and significant effect in both variable and fixed treatments, they also found that an increase of monitoring intensity over time goes along with a decrease of agents' effort levels. Moreover, when principals and agents had an opportunity to meet and discuss some personal topics ('interpersonal' condition) before engaging in the ten rounds of a partner condition, monitoring intensity negatively affected agents' performance.

#### 4.4. *Fairness and equity: ultimatum and dictator games*

As mentioned in Chapter 1 (sections 6.2 and 6.3), the emergence of cooperation in experimental social dilemma, at odds with predictions based on egoistic and myopic payoffs maximizing preferences and beliefs, leave room for various alternative behavioral assumptions. Fairness, equity and reciprocity concerns may be at stake and the distributional effect of the cooperative choice cannot always be differentiated with efficiency considerations. 'Distribution games' experiments have first been designed as simple settings to test game-theoretic predictions in the context of bargaining processes, as analyzed in Stahl (1972). In the full bargaining game two players alternatively make offers on how to divide and share a predefined amount of money or good ('the pie'). Since the size of the pie is fixed all solutions of the game other than a final rejection are Pareto optima and can thus not be differentiated by the Pareto efficiency criteria<sup>40</sup>. Furthermore, since the relative share is determined by the proposer alone, and since the game is not repeated after the end of the bargaining process, reciprocity can be controlled for as an immediate determinant of the proposers' choices.

##### *Ultimatum game experiments*

The last round of the bargaining game, when a rejection by the second player implies a failure of the bargaining process and thus no payoffs for any of the two players, has been extensively tested experimentally under the 'ultimatum game' label. Introducing the game in

<sup>40</sup> Though in these settings cooperation may thus not take the form of an attempt to reach a social optimum where *all* players are better off than in the Nash equilibrium, some contribution is needed from one player in the form of a monetary transfer to the other one in order to reduce the uneven distribution of wealth stemming from the initial move of nature. This can amount to a cooperative choice if distributional equity is considered as a collective good, which obviously depends on players' preferences. Note that the experiment presented in Chapter 5 attempts to provide a test on this very question.

experimental settings, Güth et al. (1982) used pie sizes ranging from 4 to 10 and found 50% to be both the minimum and the modal share chosen for themselves by proposers, with an average share of 65% of the pie. The other studies presented in Güth and Tietz (1990) review all find similar results, also confirmed by the findings in Kahneman et al. (1986) and Forsythe et al. (1994). This supports the assumption that distributional concerns are at stake and strongly influence players' choices, either because of the proposer's own preferences for equity or because proposers themselves widely assume that responders will exhibit such preferences. The refusal of unfair offers often observed on the second player's side in the ultimatum game decisively points to equity as an important precondition to engage in collectively benefiting social processes, confirming the proposers' beliefs assumed in the latter explanation. Using the strategy method, Kahneman et al. (1986) could determine that the minimum share considered as acceptable by responders was on average 23%, with 56% of the responders refusing shares lower or equal to 15% of the pie.

### *Dictator game experiments*

In the second experiment presented in the same paper subjects were asked to choose between two predetermined distributions amounting to a 90/10 and a 50/50 split of the pie, to their own advantage in the former case. Importantly, the receiver could not reject the offer, transforming the game into what would be later known as a 'dictator game'. 76% of the subjects divided the 20\$ evenly, giving further evidence that equity considerations are a significant component of the proposer's preferences. The second part of the experiment further showed that responders in ultimatum games may not have refused unbalanced offers only because they would suffer from the resulting payoff inequity but also as a way to punish the proposer. When asked to divide \$10 evenly with a player who had previously evenly divided the pie or to divide \$12 evenly with a player who had chosen the 90/10 distribution, 74% of the subjects chose the former, showing that they were willing to incur a loss in order to punish uneven choices and reward even choices. Forsythe et al. (1994) also present the results of a generalized version of the dictator game where the proposer could choose any distribution of the available amount, which has usually been the case in subsequent ultimatum and dictator games experiments. 36% of the players chose to keep the whole amount for themselves, which was also the modal choice. The second most chosen distribution was attributing 4\$ to the choosing player and 1\$ to the receiver, the third most chosen was the even distribution and the average amount left to the receiver was slightly above 1/5 of the pie. Considering that the uneven distribution in Kahneman et al. (1986)

amounted to a 1/10 share for the receiver, the results of the two papers are somehow converging to the conclusion that the strictly dominant selfish choice is not the majority choice though equal shares distributions are even more rarely observed. The 129 contributions reporting dictator game experiments that have been published since 1986 and reviewed in Engel (2010) altogether roughly confirm these values, with a 28% average share of the pie attributed to the receiver and 36% of the players giving nothing to the responder.

Results from dictator game experiments complete and precise the observations of split choices in the ultimatum game. They confirm that the fear of a refusal from the responder indeed partly explains the seemingly non-selfish choices made by proposers but also that considerations of equity do play a role in subjects' choices, albeit not dominant.

### **5. Asymmetric n-person interactions: three-player distribution games**

The last category of experimental games that we will review here introduce a third player in order to discriminate more finely between choice motivations. Güth and van Damme (1998) present the results of an ultimatum game experiment where the proposer had to divide the allocation between three players, himself, the responder and a dummy player, in various informational settings. In the first condition the responder was informed of the whole distribution chosen by the proposer (full information condition) whereas in the two other conditions the responder was only informed of the amount to be allocated respectively to himself (essential information condition) or to the dummy (irrelevant information condition). They crossed these conditions with two different modes, one where subjects played in the three different conditions in turn for nine rounds and another where subjects played in only one of the conditions for six rounds, so that four treatments were applied. Proposers' choices are clearly in line the results from ultimatum and dictator games, the latter being closely related to the irrelevant information condition since in both settings the proposer is protected against the responder's reaction to his own payoff as well as to their relative payoffs. Under that condition, the proposers allocated on average 75% of the amount to themselves, over all rounds and across the two modes, leaving 10% to the responder and 15% to the dummy. In the full information mode, in line with the similar ultimatum game, proposers kept on average 53% of the total amount for themselves, leaving respectively 36% and 13% to the responder and to the dummy, so that the proposer and the responder got respectively 60% and 40% of the amount shared between them. In all treatments, responders overwhelmingly accepted the offers, with rejection rates of only 7% and 8% in the cycle and in the constant

modes respectively. These observations strongly suggest that players do not only react to payoff differences but also to the perceived fairness of other players, supporting the assumptions of the intentionality models (presented in chapter 1, section 6.3), and that positive and negative differences in payoffs do not trigger inequity aversion to the same extent, for which the distinct parameters used in the Fehr and Schmidt (1999) model allow. Finally, proposers allocated on average 7% to the dummy, in both modes, when the dummy's share was unknown to the responder.

Bereby-Meyer and Niederle (2005) use a more complex design where the proposer must divide the total allocation of 10\$ either between the responder and himself (TRP – third party rejection payoff condition) or between the responder and the third party dummy (PRP – proposer rejection payoff), and where the rejection of the offer by the responder triggers a rejection payoff for the party that would have got no share of the amount in case of an agreement. These two conditions are crossed with three alternative rejection payoffs of 0\$, 5\$ and 10\$. In all three TRP treatments the proposers choices were consistent with ultimatum game observations. The 50% split appeared to be the modal choice among all possible proposed distributions, the most favorable distribution to the proposer coming second, and proposers keeping around two third of the for themselves on average. In the two PRP treatments where proposers would get a rejection payoff of respectively 5\$ and 10\$, 56% and 64% of them chose the least favorable distribution to the responder, the even split coming next with 13% in both cases. When proposers payoffs do not vary with the responder's choice, a majority of the former favored an equal share (42% of choices) while 19% chose the most defavorable distribution to the responder and another 19% chose advantageous ones. Comparing responders choices across treatments they also found the number of low offer rejections to be higher when the proposer would benefit of the offer (TRP) than when not (PRP) and to be lower when the proposer would benefit from the rejection. Both results support the intentionality models of inequity aversion since unequal offers are more likely to be rejected if they benefit their initiator even though it makes no difference in outcome in terms of payoff distribution between the three players.

The next step was to set aside strategic interaction, as was done Engelmann and Strobel (2004) who had subjects in groups of three choose between alternative distributions. Their experiment was designed in order to discriminate between several equity motives as well as group efficiency, regardless of selfish monetary preferences. Accordingly they kept the payoff to the choosing player constant in most of their treatment, varying only the maximin, the group efficiency and the two inequity aversion solutions. They found that group efficiency

concerns and maximin together could account for a large part of their observations, outperforming both Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) inequity aversion models. Moreover, their results give some clues about the self-centered feature of inequity aversion which shapes the models of fairness presented in the previous chapter (section 6.3), with players choices actually influenced by distributional concerns among other players other things equal but other-regarding equity being outweighed by other factors when the latter are at stake.

In their answer to the previous study, Bolton and Ockenfels (2006) used a similar design with groups of three subjects who had to vote for one of two alternative distributions, one which being an egalitarian split. In the straight mode, where subjects knew which role they will be matched with before voting, most of them chose the distribution which allocated them the highest payoffs when concerned. Almost one half of the subjects who would get the same payoffs in the two distributions, however, chose the egalitarian one against the alternative that would have favored one of the three players, supporting the influence of disadvantageous inequity aversion when absolute monetary stakes remain unchanged for the player.

Taken together, results from distribution games tend to indicate that equity concerns can only partially explain distributions observed in cooperation games, so that cooperative choices should be observed to a greater extent than either narrow self-interest or fairness allows of.

## 6. Conclusions

### 6.1. *Smart self-interest: far-sightedness, reciprocity and conditional cooperation*

The first conclusion that can be drawn from the above experimental results is that monetary payoffs maximizing choices are not reduced to short-term, others-ignoring behavior. Though at least some free-riding is observed in most rounds in almost all experimental settings described here and can be qualified as a strong, significant pattern of the so-called non-cooperative game situations, pursuing self-interest in the form of monetary payoffs is also at stake with those player choices exhibiting reciprocity. Reciprocity can indeed be considered as a trade-off between risk and pay-off rather than a 'non-self-interested' choice. Reciprocating players balance the risk associated with the cooperative choice – that other players do not cooperate – with the higher pay-off yielded if cooperation is achieved. The efficiency of pre-play communication in reaching the social optimum furthermore shows that



beliefs about other players' behavior are crucial in determining cooperation rates, not only tempering the possible effects of risk aversion but also in the creation of the conditions necessary for the emergence of trust.

### **6.2. *Broadened self-interest: conformity, equity and "altruism"***

Altruism, as represented by advantageous inequity aversion, which implies that subjects are willing to get lower monetary payoffs in order to decrease the payoff difference to other players, can also be detected throughout these experimental results. More generally, the four social value orientations that have been defined elsewhere (see e.g. Kuhlman and Marshello, 1975 and McClintock and Liebrand 1988) according to the possible combinations of agents concern for their own payoffs and for the payoffs to other agents - cooperative orientation focusing on joint outcome, competitive orientation pursuing the highest difference between own payoffs and others', altruism ignoring own payoffs and individualism ignoring others' payoffs - seem thus to be present in the data.

The tendency of players to imitate other players' choices that has been clearly observed in public good games (see e.g. Keser and van Winden, 2000, Bardsley and Sausgruber, 2005 or Tan and Bolle, 2007) also points to conformity as an important behavioral driver. Though conformity requires information on the choices made by others in the same or in a similar situation, it can be a decisive factor in dynamically supporting cooperation once it has emerged.

### **6.3. *Indecisiveness of collective production situations***

Taken together, all the above results constitute evidence that in the experimental world cooperation can emerge within groups of actors, without the intervention of a hierarchically empowered principal. However, in such settings the achievement of the cooperative outcome cannot be taken for granted but depends on complex interaction patterns between heterogeneous agents endowed with versatile individual preferences, as an output of highly context- and path-dependent processes.

Because of the diversity of individual behavioral patterns and of the situations at stake, no assumptions bearing one-sided conception of human behavior can be justified and no general conclusion can thus be drawn. Cooperation may emerge as a spontaneous choice of the concerned actors or will need to be elicited from some of the potential actors by other, leading or empowered, actors.

#### 6.4. *Three steps further*

Still, it is possible to draw an increasingly precise picture of the conditions under which cooperation has the highest chance to emerge, and this effort furthermore brings out valuable results since it provides ways to economize on costly mechanisms that can prove counterproductive in the long run<sup>41</sup>, or even hamper collective action and organizational learning.

The next part of this thesis presents the results of three experiments of our own. The two first experiments explore the emergence of voluntary cooperation respectively in a public good game and in a gift-exchange game. In the first one (Chapter 3) we test for the effect of the possibility for players to choose between entering the game or not. In the gift-exchange game experiment (Chapter 4), we study the share of incentive contracts relative to trust contracts chosen in a partner and in a stranger setting, and its dynamics over time.

Our third experiment (Chapter 5) focuses on the distributional concerns related with cooperation, and more precisely on the influence of the conditions under which a choice between an equitable and an efficient distributions is made. Since the cooperative choice has been explained, at least partly, by considerations of equity, it is important to finely assess the relative influence of payoff distribution, and thus to isolate distribution factors from the strategic interaction effects that can be generated by social dilemma situations, either in the form of reciprocation or through the less rationally driven non-consequential behavior. At a more exploratory level, it provides an tentative evaluation of two key elements of John Rawls' theory of justice, namely the veil of ignorance and the difference principle, and assesses the effect of procedural equity in the choice between the maximin and the pay-off dominant distribution.

---

<sup>41</sup> Though we will not elaborate on this point here, the deleterious effect of inequity, monitoring or exogenous sanctions on individual motivation, performance and cooperation has widely been illustrated.



## Collective production processes, cooperation and incentives

---

---

### PART TWO

### *THREE EXPERIMENTS*

---

---

**CHAPTER III** : *Voluntary participation in a collective-good game* .....85

**CHAPTER IV** : *Alternative designs for cooperation: incentives, trust and partners* .....109

**CHAPTER V** : *Disentangling distributional motives* .....129



---

## CHAPTER III

### VOLUNTARY PARTICIPATION IN A COLLECTIVE-GOOD GAME

---

#### Summary<sup>42</sup>

In real-world settings of collective production processes, agents seldom find themselves locked into forced cooperation situation. Since some positive levels of contribution have been observed in public-good game experiments, it is worth to examine closer individual choices, and to see what happens when players can choose to participate or not in the game.

We investigate whether the freedom to participate in the game or not can affect contribution levels over time. The analysis of two voluntary-participation treatments supports a noticeable effect of an attractive exit option on contribution levels. We conclude that voluntary participation can induce a recovery of cooperation levels when the payoff yielded by the exit option is high enough, so that the usually observed decay of average contribution levels can be counteracted.

---

<sup>42</sup> This chapter is an extended version of the paper published in the *Journal of Economic Psychology* n°31(4), pp. 705-718, co-authored with Kene Boun My (2010).

## 1. Introduction

Many tools to foster cooperation and solve the social dilemma in an efficient way have been designed and tested. However, the implementation of these tools is costly, whereas many real life examples, as well as the study of human societies, show that cooperation can be thought of as an emergent phenomenon. Thus, if free-riding behavior may lead to inefficient social outcomes, the systematic use of costly coercive mechanisms such as taxes or incentives may hinder the emergence of cooperation and reduce the associated benefits where those benefits could be non negligible (Ostrom, 2005).

We specifically address the case of collective production processes, where a sizable group of agents is generating a single outcome which benefits equally all the declared members of the group. Our case can be directly documented by the rich literature on public goods.

Three decades of public-good experiments find that almost all possible outcomes could be observed in voluntary-contribution mechanism settings, either at the individual or at the group levels. Behaviors range from pure free-riding to full contribution, depending on variations of the design and parameter setting. This high heterogeneity of choices and outcomes, combined with the subtle interaction between individual and group levels, may explain why behavioral models have failed up to now to provide decisive explanations of the emergence and persistence of cooperation. Results have mainly arisen from treatment comparisons, testing among other things the effect of group size (Isaac and Walker, 1988a, Isaac et al., 1994), multiplier value (Isaac et al., 1984), communication (Isaac and Walker, 1988b), or punishment (see, e.g., Fehr and Gächter, 1999, Masclet et al., 2003). Positive levels of contribution have always been observed in the standard settings of public good games, using the sole voluntary contribution mechanism. At the aggregate level (inter-group average) an average of 50% of the social optimum is the usual level reported for one-shot settings or first round outcomes in repeated settings, followed in the latter case by a progressive decrease (see e.g. Ledyard, 1995, for a review). We have thus some evidence that the voluntary contribution mechanism can provide a substantial provision of collective good without resorting to costly additional mechanisms such as monitoring or punishment, which may furthermore not be available. Since in many real settings where voluntary cooperation can be observed agents have some freedom to choose to interact with partners, or to leave if they are not satisfied, we have used, in addition to the standard public good game setting, a voluntary-participation treatment in which players can exit the game. This

gives us the opportunity to test the potential bias created by the standard condition in public good experiments that participation to the game is compulsory. Section 2 reviews results from seminal and recent papers related to voluntary participation and endogenous group formation in the presence of social dilemmas. Since we focus on standard linear collective good games outcomes, we refer to Ledyard (1995) comprehensive survey for this part of the related literature. Section 3 presents the experimental settings, theoretical predictions and hypothesis. We present and discuss our results in section 4 and conclude in section 6 with a summary of our results and possible directions for further exploration.

## **2. Related literature on voluntary participation**

To our knowledge, the first experimental setting implementing voluntary participation in a social dilemma game is to be found in Orbell et al. (1984). They report experimental data from one-shot  $n$ -player prisoner's dilemma games. This appears to be a hybrid form of game since it involves several players as in the standard public good game, but the choice offered to the players is binary (cooperate or defect) as in the standard prisoner's dilemma (or more precisely ternary since they added an exit option). Their main investigation was to test whether cooperators would "exit more readily than defectors", so they did not run a baseline treatment. Instead they reversed the usual order of choice and asked first the defect/cooperate choice before the enter/exit choice. Their intuition that cooperators would exit more readily than defectors was strongly contradicted by their results, since proportionally, cooperators were no more likely, and often significantly less likely, to exit than defectors. Their results further confirm the strong effect of communication in increasing cooperation levels. Their design has however some important features that reduce the significance of the results. First they do not provide benchmark results, which would have required testing the exit option against a standard game where players do not have the opportunity to exit. Second, if they actually tested two levels of the exit option named "low" and "high" exit incentives, they chose relatively high values for both parameter levels. The low exit option yields an average return equal to the expected pay-off of the game. Further, the binary cooperate/defect choice does not allow the subjects to fine tune their contribution levels. Finally, they ran one-shot games. We address all these issues: we run a baseline treatment without the exit option; we set our high exit option to the average expected pay-off of the game; we use a ten-level contribution scale; and we run a twenty-round repeated game. Subsequently, Orbell and Dawes (1993) present results from a one-shot two-player prisoner's dilemma game, with and



without the option to play the game. They observed a positive effect of the option not to play the game on the proportion of cooperative interactions.

Hauk (2003) reports results of a repeated multiple prisoner's dilemma game experiment. In the "attractive outside option" setting (i.e. when the pay-off yielded by exiting the game is higher than the mutual defection pay-off) choosing whether to play the game or not with each of the potential partner led to more cooperative relationships relative to active relationships than in the baseline treatment (without outside option) and to higher cooperation levels conditional on entry over the 10 rounds. However, a closer examination of the results mitigates this observation in the sense that the average cooperation level per round reported for the exit option treatment is conditional on entry, meaning that the overall cooperation level (percentage of cooperative relationships relative to all possible relationships) may not be higher than in the baseline treatment. In another two-player public good game, Coricelli et al. (2004) found that unidirectional partner selection has some positive effect on cooperation, though it does not prevent the decay of contributions over time.

Voluntary teaming has also been observed to increase cooperation levels in the related context of effort games and team incentives by Keser and Montmarquette (2004). Their setting involves a two-player game where each player first chooses between private and team remuneration, then chooses his effort level according to an individual cost function. They compare the outcome of this game to a baseline treatment where teaming is enforced: players cannot choose the private remuneration. They found that effort was significantly higher under the voluntary teaming treatment than under the enforced teaming treatment, in the case of symmetric effort cost functions. However, contribution levels in social dilemmas have sometimes been observed to be sensitive to group size per se (see e.g. Isaac et al., 1984 or Ledyard, 1995) and results from two-player games may not be linearly generalized to n-player games.

In n-person public-good games, closer to our setting, Ehrhart and Keser (1999) report a permanent cycle where free-riders chase high contributors, the latter continually escaping by joining smaller groups. Their results support the hypothesis of persistence of both free-riding and cooperative behavior, but the design neither controlled the net effect of endogenous group size variations, nor provided players the opportunity not to play the game.

Page et al. (2005) explored the effects of punishment and endogenous group formation. They used a standard repeated public good game setting and a  $2 \times 2$  design crossing both

parameters. Endogenous group formation was implemented by a "regrouping" treatment where players had the opportunity to give their preferences among the other 15 subjects, ranking them as prospective partners. Their results show a 70% average contribution for the regrouping treatment, as compared to 38% in the baseline treatment. Thus, introducing some freedom in the constitution of groups already increased the contribution level. However, this did not appear to be enough to trigger sustained contribution, the results showing a clear decay through the 10 rounds. In addition, the experimental design relies on the possibility of an immediate regrouping (groups are rebuilt from one round to the next), which is seldom the case in real settings. Finally, as in Ehrhart and Keser (1999), players cannot exit and stop playing the game.

All these results begin to shed some light on the positive effect of voluntary participation on cooperation levels. However, this stream of work has several limitations. First it is mainly concerned with two-person settings. Second, it does not provide clear clues on the effect of an effective exit option at the aggregate group level. Last, it does not established whether voluntary participation is an efficient feature for sustaining contribution levels when players can really choose not to play the game.

In order to test the robustness of the previous findings in a context of repeated interactions within groups of agents, we extend the settings to an n-player repeated game where the individual pay-off depends on the level of cooperation within the participating group.

### **3. Experimental settings and treatments**

The experiment was carried out at the Laboratoire d'Economie Expérimentale de Strasbourg (LEES) using using our in-house designed data-processing software. 135 voluntary subjects took part in the experiment after being randomly selected through ORSEE<sup>43</sup> among 1200 students from various programs. None of them had previously confronted public good experiments (inexperienced subjects). Written instructions were distributed and read aloud to the subjects<sup>44</sup> before they performed a pre-experimental test to check proper understanding. The session began as soon as they all had correctly answered every question. No communication was allowed between subjects as long as the experiment was running. We applied three treatments, namely B, VP-11 and VP-16 (B for baseline treatment, VP-11 for

---

<sup>43</sup> A web-based Online Recruitment System for Economic Experiments developed by Ben Greiner.

<sup>44</sup> See Appendix A for the original instructions (in French).

voluntary participation treatment with a fixed exit payoff of eleven and VP-16 for voluntary participation treatment with a fixed exit payoff of sixteen).

For each treatment we formed randomly nine independent groups of five subjects who played twenty rounds with the same partners. Each subject participated in one treatment only and in only one group, so that group data are independent from any group to any other. Table 3.1 summarizes the experimental treatments.

*Table 3.1: Experimental design.*

Treatment	Number of sessions	Number of independent groups	Number of subjects
B	3	9	45
VP-11	3	9	45
VP-16	3	9	45
Total	9	27	135

### **3.1. The baseline (B) treatment**

In order to link our experiment as closely as possible to existing work either using a voluntary participation mechanism or examining other determinants of cooperation in a traditional public good game setting, we used the most standard form of linear public good game as baseline treatment. This furthermore gave us some clues on what values to choose for the parameters since some combined effects of group size and multiplier value in public good games have already been investigated. Though the net effect of group size variations is not always precisely predictable, it has been shown that for reasonably high values of the multiplier (private return of the collective investment) contribution levels are not correlated with group size (see Isaac et al., 1984 and Isaac et al., 1994).

We formed random groups of five subjects and kept the same groups for a repeated game of twenty rounds (since we focus on cases where the identity of the agents determines the potential interactions we only used partner treatments). At the beginning of each round, each player  $i$  is endowed with  $Y_i = 10$  tokens, and chooses how to distribute them between a private account and a public account, named "project". Each player has thus to split her endowment between the two investment possibilities so that all her tokens are invested. The

profit  $\Pi_i$  of a player is determined by her contributed amount  $C_i$  and by the sum of the contributed amounts of the  $j$  other players, and is defined as:

$$\Pi_i = Y_i - C_i + \alpha \sum_{j \neq i} C_j \quad (1)$$

with  $\alpha = A/N$ ,  $0 < \alpha < 1$ ,  $A > 1$  and where  $\Pi_i$  is player  $i$ 's pay-off,  $Y_i$  is the individual endowment.  $C_i$  is player  $i$ 's contributed amount to the collective account and  $C_j$  is the individual contribution of the other participating players. We set  $\alpha$  at 0.75,  $Y_i = 10$  and used nine groups of five players for each treatment.

Setting the parameter values for the baseline treatment is a crucial choice since the number of players involved in the public good game varies in the second treatment. In order to set an acceptable basis for comparison it was thus necessary to discuss various possible ways of declining expression (1). Varying the number of players in a public good game is not trivial, and to date no clear answer has been given to the effect of group size per se on cooperation levels. In the voluntary participation (VP) setting, the potential total contribution increases with the number of participating players if  $\alpha$  is kept constant. Hamburger et al. (1975) and Bonacich et al. (1976) give some clues on what can be considered as equivalent public good games when group size varies, which led us to keep  $\alpha$  constant so that the marginal substitution rate from the private to the public account is unchanged whatever the number of players participating in the game ( $N$ ). This also means that the potential individual payoff increases with group size, which can be interpreted as the effect of increasing marginal returns of the collective production function within the scale of the group size (from two to five).

The total contribution of the group is common knowledge at the end of each round and players can display the history of their own contributions and of all past total contributions of the other players. The composition of the group is the same all along the game, and players do not know the identity of their counterparts.

### *Theoretical predictions and hypothesis*

The standard game theoretical outcome of the linear public good game we use here is well known to be that no contribution will be made to the public account by self-interested, rational players who have correct information, since contributing nothing is the only pure dominant strategy for every player. The only Nash equilibrium is thus that all players free-ride and none contributes. This result holds in both the one-shot and in the repeated games. However it has been observed and is considered as a consensual result that contributions will

amount around 50% of the social optimum in the first rounds of repeated treatments, and then steadily decrease without reaching a null value. This outcome can be explained by a number of alternative theoretical assumptions or models that cannot be fully described here (fairness, reciprocity, conditional cooperation, learning).

Our first hypothesis is that *positive and decreasing average contribution levels will be observed in the baseline treatment.*

### 3.2. *The voluntary participation (VP) treatments*

In order to test for the direct effect of voluntary participation on contribution levels, we inserted in the two other treatments a first step before the public good game was played, namely the opportunity for players to choose to participate or not in the public-good game. We thus depart from the principles of pure public good that underlay the voluntary contribution mechanism, in order to integrate at the same time two properties of group projects: only identified participants of the project get some reward from the profits associated therewith, which means nothing other than attributing a perimeter to the non-excludability of the "good" ("good" standing for the value created by the agents). Conversely, subjects have thus the opportunity to isolate themselves from the consequences of others' behavior by opting out. In that case, they get a fixed income, cannot benefit from the collective project but also cannot incur losses induced by other players' choices. The difference between the two VP treatments is the payoff associated with the exit option. In the first voluntary participation treatment (VP-11) the pay-off associated with the exit option is the minimum value higher than the minimum certain payoff procured by the collective-good game, so that exiting is a strictly dominating strategy of the stage game. In the second voluntary participation treatment (VP-16), the exit payoff is the closest value to the average expected payoff of the collective-good game.

We use two different payoffs for the exit option in order to control for the effect of the possibility to choose between participating in the game or not, and to test if a more attractive exit option could be more effective as a credible threat. Indeed, in the latter case, it is clear that contributors have a secure alternative in opting out and getting the exit option payoff. The income per round is thus determined as follows: if a player chooses to opt out, then he gets a fixed income (11 in the VP-11 treatment or 16 in the VP-16 treatment); if a player chooses to participate in the collective project, he then plays the public good game with the other participants if there are any, or gets a payoff of 10 if he is the only participant.

Before choosing their contribution players participating in the collective project know how many people in their group have chosen to participate. The number of participating players and the total amount contributed to the collective project is common knowledge after the end of the round. Since the effect of group size on contribution levels is not clearly established, we first planned to apply a  $2 \times 2$  treatment design to test two different pay-off functions, one where  $\alpha$  was hold constant whatever  $N$ , and a second one where  $\alpha$  was varying against  $N$ . However, the second function implies a possibly substantial variation of the marginal private incentive not to contribute if the number of players choosing to opt out varies and implies a strong constraint for the value of the multiplier if all players participate.

### *Theoretical predictions and hypothesis*

As for the two voluntary participation treatments, they should theoretically display no difference in the outcome: in both cases the payoff associated with the exit option is such that opting out is a strictly dominant strategy, entailing a unique Nash equilibrium where no player enters the game. However, since we expect substantial contribution levels in the first rounds of the linear game, we also expect players to choose to enter the game and contribute. Once in the game, the situation can be seen as equivalent to the baseline treatment plus the exit option as a credible threat.

Since voluntary participation has already been shown to increase cooperation levels, we furthermore expect contribution levels in the VP treatments to be higher than in the Baseline treatment.

Our second hypothesis is that *an attractive exit option has a positive effect on the variation of contribution levels over time*, either in giving risk-averse players an opportunity to exit the game or in providing players with a credible threat.

## **4. Results and discussion**

We collected and analyzed for each treatment the average contribution to the public good, in percentage of the socially efficient outcome ( $C_{i,j} = N * Y_i = 100\%$ ), and the number of participants. For the voluntary participation treatments we computed the average contribution both for the whole group ( $N = 5$ ) and conditional on entry ( $N =$  number of players opting in).

**4.1. Choices heterogeneity: neither Nash nor the optimal social outcome**

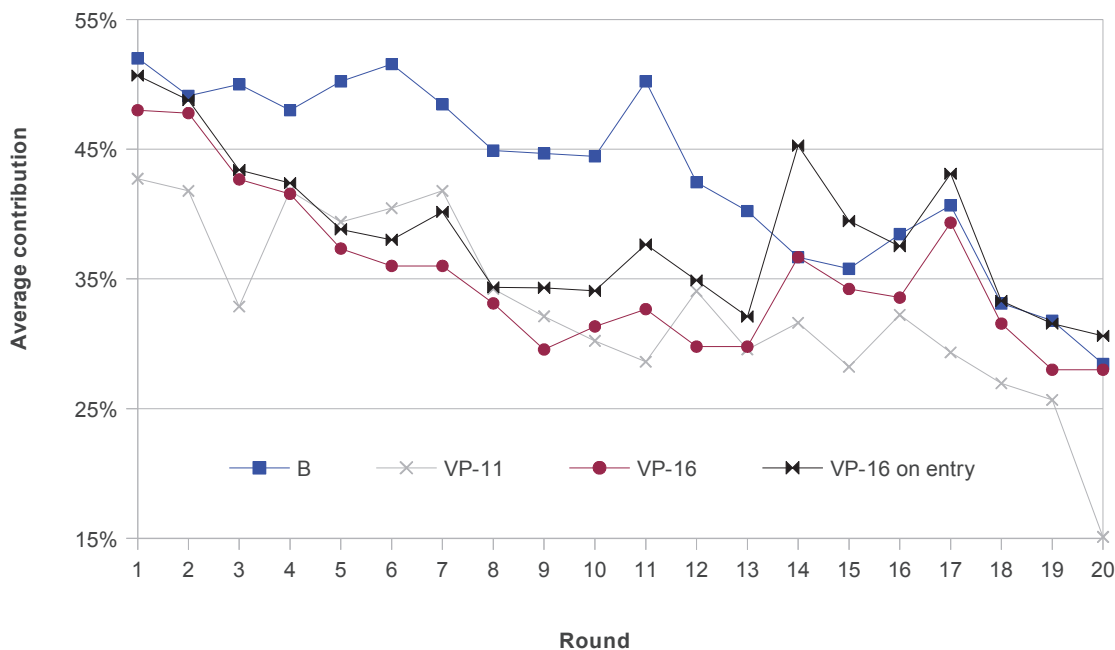
Table 3.2 displays the average contributed amount in percentage of the social optimum outcome for each treatment, averaged on the twenty rounds. The values of average contribution levels among the nine groups of each treatment and for each round are displayed

*Table 3.2: Averaged contributions per treatment.*

Treatment	B	VP-11	VP-16	VP-16 (on entry)
Average contribution	43%	33%	35%	39%

in Figure 3.1. Our aggregate data follow the usual pattern observed in public good game experiments (see Isaac et al., 1984) with a positive contribution to the collective project in the first round, around half of total players endowment, followed by a significant decay.

*Figure 3.1: Temporal dynamics of the average contributions, per treatment.*



In the three treatments we also observed a wide variety of individual sequential choices. They range from pure free-riding<sup>45</sup> (around 20% of cases in each treatment, around 10% of

<sup>45</sup> Following Isaac, Walker, and Thomas (1984) and Burlando and Guala (2005) definitions of free-riding, we only use the terms 'free-riders', 'free-riding' or 'pure free-riders' to denote players who contributed nothing since the very first round of the game.

**Table 3.3:** Round-to-round adjustment of contribution level according to the situation of the player's contribution relative to the average contribution of the other players in his group, per treatment and per situation, in share of total corresponding observations.

(number of cases)

Situation*	Treatment	Increase	Decrease	No change
1	B	19% (78)	55% (223)	26% (105)
	VP-11	21% (68)	50% (163)	30% (98)
	VP-16	17% (52)	55% (167)	28% (94)
	All	19% (198)	53% (553)	28% (105)
2	B	48% (200)	16% (67)	36% (149)
	VP-11	32% (61)	14% (140)	54% (235)
	VP-16	47% (152)	17% (56)	36% (117)
	All	35% (413)	22% (263)	43% (501)
3	B	40% (12)	13% (4)	47% (14)
	VP-11	34% (12)	11% (4)	54% (19)
	VP-16	35% (7)	40% (8)	25% (5)
	All	36% (31)	19% (16)	45% (38)
All		28% (642)	36% (832)	36% (836)

\* Situation 1: Player's contribution higher than average contribution of other players

Situation 2: Player's contribution lower than average contribution of other players

Situation 3: Player's contribution equals average contribution of other players

players for treatment B and VP-11, and 6% of players for treatment VP-16 including free-riders who eventually exited) to full contribution (around 10% of all cases and 8% of all players choosing full contribution most of the time), and from unit-by-unit fluctuations to nothing-to-all alternations. We can notice here that in both VP treatments, subjects overwhelmingly failed to recognize that it was a dominant strategy to exit: the exit option was chosen only in 3.5% of all cases in VP-11 and 16% in VP-16.

#### 4.2. Observed average behavior as a strong attractor

As in Ashley et al. (2002) we found a positive correlation between player's first round contribution and their average contribution in the other rounds (Pearson's rho: 0.62). In line with Keser and van Winden (2000) we could also detect a tendency of players to adjust their contributions in the direction of the average contribution of the other players (see Table 3.3). Almost no random behavior is identifiable and groups exhibit strong homogeneity in



contributions levels over time, with higher contributing groups tending to keep a high and more sustained average contribution level.

The first available direct empirical evidence of social influence in public good games is given in Fischbacher and al. (2001). Using the strategy method they found that 64% of the subjects would choose their contribution levels according to the average contribution levels of the other members of their groups. In their experiment, however, subjects were explicitly asked to indicate their choices relative to the average contributions of the other players in their group, which can be seen as a source of framing effects, and they played only once. The whole dynamics of collective processes is thus not captured and can not be directly inferred from static individual observations.

We can get a first clue on individual reactions to the observable average behaviors of the other players using the frequencies displayed in Table 3.3. We can see that the propensity of players to change their contribution level in the direction of the average contributions of the other players of their group is significantly high. In all three treatments, players tended to adjust their choices towards the observed average contribution levels. Interestingly, Keser and van Winden (2001) found very similar proportions in their partner setting (4 players, 25 rounds). Up to now we have detected a net tendency of subjects to move their contribution levels in one direction, and we want to assess whether the amplitude of these moves is related to the quantitative difference between subjects' choices and the average choices of their counterparts. We computed the differences between previous round average contributions of the other players ( $LagAvC_j$ ) and the previous round contributions of player  $i$  ( $LagC_i$ ), and the difference between player  $i$ 's present contributions ( $C_i$ ) and previous round contributions ( $LagC_i$ ). There happens to be a significant correlation between situations relative to the average contribution of the other players of the group ( $Diff = LagAvC_j - LagC_i$ ) and variations of the contribution level ( $Var = C_i - LagC_i$ ). The Spearman rank correlation test reveals a strong positive relationship between  $Diff$  and  $Var$  ( $\rho = 0.60$  for treatment B,  $0.54$  for treatment VP-11 and  $0.64$  for treatment VP-16, all significant at the  $0.001$  level). These findings are all the more significant than we did not exclude any player from the dataset. This means that the correlation values hold for a heterogeneous panel also including the contributions of players exhibiting the specific strategies of free-riding, alternating high/low contribution, or even random choice, and hold for all groups. It clearly supports the assumption that players tend to adjust their contributed amount to the average practices observed in their group, in all three treatments. A similar effect is observed by Tan and Bolle (2007) at the inter-group level, in a setting where players were informed if their

group's total contributions had been higher or lower than another group at the previous round, and in Bardsley and Sausgruber (2005) where players are informed either of their own group contributions or of non relevant group contributions, or both. Their results show that subjects respond to inter-group comparisons in the direction of the observed behavior of the other players.

These observations can be directly related to the social psychology concepts of social proof (Cialdini, 1993, Macy, 1993) and, though somehow more remotely, of cognitive dissonance (Festinger, 1957). The latter is at stake whenever a conflict arises between cognitive elements in a subject, for instance between beliefs and perceptions. In our case, instead of applying a predetermined strategy possibly including some adaptation to the observed play of other players, subjects would tend to adapt their beliefs – their perception of the game – according to the observed average behavior of their counterparts. With some inertia, though and some exceptions. The former, also named informational social influence, accounts for one possible way of managing uncertainty in situations where people are unable to determine the appropriate mode of behavior, namely the tendency of actors to imitate those around them. In the context of public good games, experimentation, which is acknowledged to be a major mode of learning, needs thus to be cautiously weighed before being integrated in models of individual decision making since it may well be bounded by the influence of emerging collective norms.

Further investigation at the group level may give some finer indications on player types and on interactions between players according to their type. The reaction to other players' behavior may also be influenced by the way it is observed. In our setting, players knew the history of the average contributions of their counterparts and not the vectors of contributions.

Although the correlation results seem to indicate that group average contributions would to some extent stick to their initial values, the latter working as attractor to which individual contributions converge, this would be in contradiction with the decreasing trend of the contribution levels usually observed in the aggregate level (averaged over all groups) even excluding the end-effect. Three observations can explain why this is not the case. First, as could be expected, the individual contribution levels are not fully flexible to the observed average contributions of the other players. A strong positive correlation has already been found between the individual contributions in the first round and the average individual contributions over all rounds (Ashley et al. 2003). Second, we can indeed see that the propensity to keep the contributed amounts unchanged is significantly higher for players whose contribution levels was beneath the average contributions of the other players in their

groups (Table 3.3, situation 2). Third, 'over-contributing' players (relative to the average contributions of the other players) tend to decrease their contribution levels more than 'under-contributing' players tend to increase theirs. This last observation can be brought together with experimental results showing that inequality aversion tends to be asymmetric in the sense of a greater sensibility when the outcome is seen as more favorable to others<sup>46</sup>.

Looking again at the aggregated data we cannot find strong support to the conclusion that contribution levels tend to decay inevitably, but rather confirmation of the behavioral assumptions supporting that contribution levels stabilize and fluctuate around some positive values, excepted in groups where one could expect that free-riders are the majority. A closer observation of the evolution of average contribution levels indicate no clear decay tendency between the initial adjustment and the very last rounds end-effect (see Figure 3.1 here, or other results from similar settings, for instance in Isaac et al., 1984, p.134-135, in Fehr and Gächter, 2004, p.187, Keser and van Winden, 2000, p.27, Tan and Bolle, 2007, p.136, and Isaac et al., 1994, for various numbers of players and rounds).

Taken together, these results support the hypothesis of a permanent adjustment of individual contribution choices towards the fluctuating average contributions of the other members of the group. More precisely, individual contribution choices can be described as the outcome of the combined attraction of autonomously determined contribution levels (i.e. the contribution levels which they would choose according to their own perception of the game) and observed average contribution levels of the other members of the group, with the relative importance of these two components varying among players. Thus, the highest the homogeneity of group members in terms of autonomously determined contribution, the highest the probability of sustained contribution levels.

This conclusion is still compatible with the distribution of 'reaction functions' observed by Fischbacher et al. (2001). They found that 50% of the subjects were conditional cooperators, in that the contribution levels they would choose is increasing with the average contribution levels of the other players. Most of these conditional cooperators nonetheless exhibit a "self-serving bias" in the sense that they would contribute somehow less than what the others contribute on average. Around one third of the subjects are strategic free-riders. The general dynamics observed in almost all collective-good experiments can thus directly be derived from their results: "since subjects react on average conditionally cooperatively on other subjects' contributions (but with bias in the selfish direction), positive but deteriorating contributions to the public good are observed". As mentioned by the authors, however, the

---

<sup>46</sup> On this point see for instance the findings in Charness and Rabin (2002).

variant of the strategy method that they used to elicit subjects' preferences implies "the assumption that elicited preferences are stable (i.e. the assumption that [the reaction function] does not change with experience)". Our results indeed support the argument that once in the repeated game situation players get exposed to the real behaviors of their counterparts, here for twenty rounds, and not any more to potential average contribution levels. In other words, social influence is stronger and individual preferences more flexible when subjects are playing the dynamic real game than when the strategy method is used to elicit their preferences. This could explain why we observed fewer pure free-riders (around 10% of the players, all treatments together) than they did but a fairly higher proportion if we include those who eventually contributed some positive amount, possibly only for one or a few rounds (around 20%) and a share even closer to theirs if we follow Isaac et al. (1984) in defining strong free-riding as accounted for by contributions between nothing and one third of the individual endowment (36% of all players). Playing the game and getting into the interaction with other players seems to create by itself a softening effect on free-riding choices. As for conditional cooperators, the "self-serving bias" observed by Fischbacher et al. (2001) may not be strong enough to generate an irreversible decay process in each and every cases, or may be corrected by the mere repetition of the game. When multiple rounds are played, players know that their own choices will contribute to some extent to the future decisions of the other members of their group and may thus not behave myopically.

Our results further highlight that contribution levels in the first rounds are crucial for sustaining a substantial investment in the collective project, and that the relative heterogeneity among players belonging to the same group in terms of perception of the game, self-determined intended behavior and expectations is a key factor in the emergence of cooperation.

### **4.3. Treatments VP vs. treatment B**

#### *Average contribution levels*

At every round, treatment B exhibits a higher average contribution level than treatment VP-11, though the Wilcoxon rank-sum test for group contributions averaged over all rounds shows only weak evidence (p-value of 0.17), so that the persisting difference can be imputed to uncontrolled heterogeneity of the players. To be sure, a low exit option does not have any positive effect on contribution levels.

One possible explanation is that the exit option can induce players to behave more strategically, as if entering the game was associated with an opportunity cost (i.e. losing the fixed pay-off from opting out). This would be an undesirable effect of our VP settings, creating a bias towards lower contributions. It is worth noticing that the exit option was chosen only in 3.5% of all cases in VP-11, showing that players mostly failed to recognize or did not play the Nash equilibrium choice when the payoff associated with the exit option was 11. At that level the exit option yields only 1 (10%) more than the Nash equilibrium of the public good game.

As for treatment VP-16, no significant difference with treatment B can be found. P-values for the Wilcoxon rank-sum test are 0.4 if VP-16 average group contribution values are computed for the whole group, with  $N = 5$ , and 0.56 if the values are computed conditional on entry. Here again, players largely failed to recognize or did not play the dominant strategy (16% of all cases), even though the payoff associated with the exit option was substantially higher than the Nash equilibrium of the public good game.

#### *Contribution rates over time*

Though we did not find any significant difference between the baseline and voluntary participation treatments in terms of average group contributions, the shape of the VP-16 contribution levels series let one think that the usual decay observed in public good games was somehow contradicted in this setting. In order to get some clue about the possible difference between treatments concerning the variation over time of contribution levels we computed chronologically symmetric average contribution levels.

We first compared contribution levels averaged over the ten first rounds and the average over the ten last rounds, for each treatment. In order to take into account the potential bias generated by the end-effect we repeated the same computation using shorter time spans. Indeed, the final-round tendency towards zero contribution usually observed in public good experiments triggers a general decrease of contributions, which can be solely attributed to the finite horizon of the game. We are rather interested by settings where a possible end of the game for a player is to exit, but where the game can be infinitely repeated. We identified the end-effect rounds according to the variations of the proportion of players choosing the dominant strategy. From round 17 on we observed a sharp and constant increase of the number of players choosing to contribute nothing in the Baseline, and of the number of players opting out in VP-16 (from 8% in round 17 up to 22% in round 18). Accordingly, we computed adjusted variation rates excluding the three last rounds of the time series. We also

excluded the three first rounds so as to keep time spans symmetric and to leave first round errors out of the analysis. The variation rates of average contribution levels are presented in Table 3.4. The first and the last rounds of the series included in the computation are indexed on ER.

*Table 3.4: Variation over time of contribution levels by treatment.*

Treatment	ER <sub>1-20</sub>	ER <sub>4-17</sub>
B	-22%	-14%
VP-11	-25%	-18%
VP-16	-16%	-4%
VP-16 (on entry)	-10%	3%

When all rounds are included in the computation, we can see that the rate of decrease of average contributions conditional on entry in treatment VP-16 is less than half than the decrease rate observed in the Baseline treatment. In absolute value the difference is equivalent to less than one token per player in VP-16, as compared to around two tokens per player in Baseline. The difference between the two treatments is weaker if the computation for VP-16 contribution levels is made relative to the whole group, but the decay is still more limited in VP-16 than in Baseline. VP-11 exhibits the strongest decrease rate, suggesting that voluntary participation per se is not sufficient to counteract the general decay of cooperation in public good games.

If the time-span is reduced in such a way as not to include the end-effect (i.e. if variation rates are computed excluding the three first and the three final rounds), the variation rates of the three treatments are ranked in the same order as previously but the difference between VP-16 and Baseline is sharper. In addition, the variation for VP-16 is positive if contribution levels are computed conditional on entry. These observations are confirmed by Spearman's rank correlation between rounds and treatment average contributions ( $\rho = -0.89$  for B and  $\rho = -0.62$  for VP-16 when the test is run over all rounds, and respectively  $\rho = -0.82$  and  $\rho = 0.03$  when the test is run over rounds 4 to 17).

Using Page's trend test on average contributions per treatment per round, we found similarly that the null hypothesis of the absence of a decrease trend can be rejected at the 0.05 level for

all three treatments when all rounds are considered. When the end-effect is isolated, the result of the test is the same for treatment B and VP-11 but the null hypothesis cannot be rejected for treatment VP-16 (p-value: 0.368).

These findings support our hypothesis that an attractive exit option has a positive effect on the variation of contribution levels since no tendency towards decreasing contribution rates can be identified in VP-16 average contribution conditional on entry when the end-effect is isolated.

Table 3.5 displays a finer description of the contribution rates dynamics observed in each treatment. We divided the whole time span into four blocks of five rounds and computed for each block the variation rate, using the first round in the block as a benchmark. The first and the last rounds of the corresponding time block are indexed on ER. As we will discuss below (see 4.2.3), VP-16 exhibits a much drastic adjustment during the first five rounds than B. Though variation rates are quite similar in the second time block, the subsequent reversal of the contribution trend observed for VP-16 in the third block contrasts sharply with the highest decreasing rate observed for B. Finally, the end effect seems stronger in B than in VP-16, though no significant difference can be found.

*Table 3.5: Variation of contribution levels within time blocks, by treatment.*

Treatment	ER <sub>1-5</sub>	ER <sub>6-10</sub>	ER <sub>11-15</sub>	ER <sub>16-20</sub>
B	-3%	-14%	-29%	-26%
VP-11	-8%	-25%	-1%	-53%
VP-16	-22%	-13%	+5%	-17%
VP-16 on entry	-23%	-10%	+5%	-18%

### *Group dynamics*

Since the differences we found between treatments could have been generated by one or few extreme group values we applied the same procedure to the average group data and ran Wilcoxon rank-sum tests to compare each VP treatment with the baseline (Table 3.6). We can immediately see that none of the nine groups under the Baseline treatment exhibits positive variation rates between the first and the second time series. By contrast, in treatment VP-16, the variation is positive for three groups whatever the time span used.

Table 3.6: *Variation of contribution levels by treatment and by group.*

Treatment B									
Group	1	2	3	4	5	6	7	8	9
ER <sub>1-20</sub>	-22%	-22%	-51%	-32%	-2%	-33%	-11%	-15%	-26%
ER <sub>4-17</sub>	-8%	-21%	-35%	-24%	0%	-29%	-5%	-13%	-9%
Treatment VP-11									
Group	10	11	12	13	14	15	16	17	18
ER <sub>1-20</sub>	-13%	-73%	-15%	-45%	-11%	-62%	-64%	14%	-32%
ER <sub>4-17</sub>	-8%	-73%	2%	-27%	1%	-44%	-66%	18%	-29%
Treatment VP-16									
Group	19	20	21	22	23	24	25	26	27
ER <sub>1-20</sub>	-44%	-64%	-25%	-39%	4%	2%	-67%	-15%	58%
ER <sub>4-17</sub>	-26%	-26%	-1%	-34%	11%	1%	-48%	-10%	68%
Treatment VP-16 on entry									
Group	19	20	21	22	23	24	25	26	27
ER <sub>1-20</sub>	-41%	-45%	-4%	-15%	2%	-1%	-47%	-14%	60%
ER <sub>4-17</sub>	-25%	-2%	27%	-7%	11%	-1%	-31%	-8%	71%

Within the VP-11 groups, the group with highest contribution increases its average contribution level by 14% and two others exhibit a positive though weak variation rate if the time-span is reduced. This can be interpreted as more evidence that the exit option is not a neutral feature and can even, at the group level, trigger a reversal of the 'downwards spiral' of declining contributions usually observed in public good experiments.

If we compare the number of groups where average contribution either increases or decreases by less than one token (variation rate between -10% and 0%), there is only one out of nine in Baseline as compared to four out of nine in VP-16 conditional on entry with the ten-round time span, and respectively four (less than half) and seven (more than half) with the seven-



round time-span. This tends to support further the idea that voluntary participation can sustain contribution levels in such a way that they may even increase over time.

We further found that the variation rates over time of group contributions were significantly higher in treatment VP-16 than in treatment B (Wilcoxon rank-sum test), rejecting the null hypothesis that the two treatments showed no difference on the variation rates of group contributions with a p-value of 0.08. We did not find a significant difference between B and VP-11, in either time span.

We also observed over all three treatments that in the highest contributing groups, contribution levels tended to stabilize more than in other groups. Running Page's trend tests on group average contributions we found that the groups for which the null hypothesis of no decreasing trend could not be rejected also exhibited the highest average contributions over the whole time span. We obtained the same result for all treatments. Among those groups, in the ones in the voluntary participation treatment, the average contribution and the homogeneity among contributions were thus high enough through all rounds to dominate the choice to opt out. By contrast, and expectedly, players in lowest contribution groups opted out much faster in VP-16 than in VP-11, and much more (16% vs. 3.5% of all choices). This can also be one of the reason why the average contribution level in treatment VP-16 follows a more decreasing path than in treatment Baseline at the beginning of the game: players exit non-profitable collective-good games as soon as the exit option is more attractive, instead of progressively decreasing their contribution as players in the Baseline treatment usually do. Conversely, since pulsing is the only available mean for players to influence others' contributions in treatment Baseline, it creates an incentive to increase their own contribution that would be ruled out by the exit choice in treatment VP-16.

*Individual behavior: the exit option as a (non-)credible threat?*

As for the net effect of the exit option on players' contribution choices, closer observation is needed. In the VP-16 treatment, the average contribution level over all nine groups is similar to the Baseline treatment for the two first rounds. It then decreases rapidly down to the level observed in the VP-11 treatment, up to round 13 where it visibly increases and reaches again the level displayed by the Baseline treatment. In between, the difference between the two treatments is only marginally significant, and from round 14 onwards the difference between average contribution levels is negligible.

Looking at the individual data of treatment VP-16, we could observe one recurrent pattern of the use of the exit option that differs from VP-11 and that could explain the unusual shape of the average contribution level series in VP-16 : first some higher contributors (defined as the players exhibiting average contributions to the collective project since the first round higher than group average, HC, as opposed to low contributors, LC ) exit, leaving the others in a situation very similar to the "low contribution" groups where contribution is still decaying. Contrarily to what happens in VP-11 settings, though, the exit option becomes immediately highly attractive and the remaining players either exit or increase their contribution to the collective project. For some reason (LC may observe that they had been much better off before opting out, HC may notice that total contribution level is increasing again) exiting players join back. This dynamic process fully occurred in groups exhibiting middle contribution levels, which could be described as 'mixed population' groups and represent one third of our observations in VP-16 treatment. In those groups, thus, the voluntary participation mechanism did not properly drive the LC out of the project, but rather induced them to join back and contribute higher than their "natural" level ("natural" meaning corresponding to their level before or without the exit option). Since only a rough indication of other players' individual profits is available to players, as long as HC are better off in the project LC (and among them pure free-riders) are induced to reproduce their choice. From the point when the HC 's interest is to opt out, the LC are faced with the choice of increasing their contribution (pulsing or signalling) or opting out too.

The outcome is a noticeable and durable impulse on the average contribution level in those groups, which can clearly be seen at the aggregated inter-group average too (see Fig. 1). In the VP-11 treatment, the exit option was too low to be used to create this kind of credible threat. The described dynamics do not directly support the hypothesis of the use of the exit option as a systematic signalling or punishment tool, and do not imply a net positive impact of voluntary participation on inter-group average contribution levels. However, it has been shown that the existence of an attractive exit option could trigger a reversal of the decaying tendency of contribution levels in groups where free-riding behavior would have otherwise undermined cooperation.

## **5. Comments**

Before concluding we would like to mention a few points that we think need to be kept in mind for an appropriate interpretation of our results. First, concerning the return to

contributions, which can be considered as a high one: since our design implied variations in the size of the group playing the public good game, we used a high multiplier value in order to avoid correlation between group size and contribution levels. Furthermore, though a 0.75 return makes indeed the public good rather attractive it has been observed (and confirmed in our Baseline treatment) that it is not high enough to generate sustained contribution levels over the repeated game. In addition, the attractiveness of the multiplier must be assessed relatively to the value of the exit option payoff, including the 'high' one (in treatment VP-16) which mechanically made playing the game less attractive.

Second, it is worth noting that the different treatments trigger difference in strategic behaviors : many behavioral models could explain our results but only tentative explanations can be inferred from experimental data. We are thinking to address this point by assessing the relative performance of alternative models using parameter estimation methods but this would be a case for a separate study with a different focus. We already discussed in this paper how some players in VP-16 did not increase their contribution level until the exit option was used by other players in the group, possibly in order to get them back in the game, and how players entering the game could have higher expectations about others' contributions as compared to a setting where opting out is not possible. Unfortunately we could not control for treatment effects in terms of strategic behavior since each subject played only one of the three treatments. It may be suggested to design a follow-up study including a crossed treatment in order to track changes in strategic behavior. VP games are indeed cognitively more complicated than the B game. In our view the main possible implication is that players may have opted out as a default choice ensuring them a certain payoff and allowing them to skip the uncertainty pertaining to the contribution choice. However our results bare no evidence that this pattern had much influence on play since exit choices amounted to only 16% of all choices in VP-16 and 3.5% in VP-11.

Third, concerning the contributions to the collective good from the group as a whole: introducing an exit option in the way we implemented it did not increase the contributions to the public good from the group as a whole, if one considers the full group as being the 5-player group. It has been observed however that in the Baseline treatment none of the nine groups exhibits positive evolution rates whereas in VP-16 both 'self-selected groups' and 'whole groups' calculations show that some groups actually increased their contributions, at least at some points within the whole time span. It is important to remind here that the most dramatic decreases have also been observed in VP-16 groups. As discussed earlier at the end of Section 4.2, the exit option may thus have had a screening effect, the opportunity to exit

accelerating the decreasing rate of contributions in low contribution groups, resulting in lower contributions from the group as a whole. Furthermore, though providing an outside option can be considered as coming at a cost for the organization it can also be viewed as the real-world possibility to engage in autonomous, independent activity. In the latter case, the loner also contributes to the organizational output though some of the potential collective benefits of cooperation may be lost. Of course a group of exiting players yields less to the collective output than any group where at least one player contributes one unit, hence the need to test the potential benefits of the exit option in terms of contribution levels. All in all social welfare implications of our findings are quite indecisive. However our focus is rather at the group level, and our conclusions are accordingly drawn less in terms of 'society as whole' than in terms of mechanism design for the provision of collective goods involving small groups or for group production processes. 'Society' in the VP treatment should thus be considered as limited to the players entering the games: exiting the game equals exiting society, hence our focus on the VP16 results on entry. As we had been thinking about during the experiment design phase, and as it has been often suggested, one could imagine a VP design where group size would be kept constant throughout the game. This would however not come without drawbacks: groups of less than five players would certainly be formed unless some subjects would be prevented from playing the game or some players would be somehow simulated and, most problematic, group consistency could not be granted.

## 6. Conclusion

Our results provide evidence that the freedom to participate or not in public good games has some positive effect on players' contributions over time and can counteract the decay of contribution levels usually observed. More precisely, a positive variation of contribution levels was observed in the third part of the whole time span in treatment VP-16 where an attractive exit option was available. It also appears that the possible alternatives need to be attractive enough compared to the potential payoffs associated with participating in the game, otherwise the exit option has no positive effect.

Two kinds of implications follow. First, though the exit option *per se* was not enough to reach the social optimum contribution levels, it triggered a revival of cooperation levels without resorting to coercive mechanisms. This gives a strong indication that emerging cooperation can be supported by the mere dynamics of group formation and players' voluntary participation thereto. More generally, the voluntary contribution mechanism

should thus be considered as a possible candidate for the provision of collective goods involving small groups, and in a way as a first candidate to be tested before the implementation of costly coercive mechanisms. The latter may furthermore hamper trust and motivation.

Second, considering the large range covered by public good situations where agents can effectively choose to enter the game or to exit, experimental settings where participation to the public good game is compulsory may generate an important bias. Moreover, the heterogeneity among players that explains inter-group differences is not visible in the aggregated data, which calls for a careful examination of the interplay between treatment effects and group effects in experimental studies since treatment effects may not be homogeneous across groups.

---

## CHAPTER IV

### ALTERNATIVE DESIGNS FOR COOPERATION

#### INCENTIVES, TRUST AND PARTNERS IN A GIFT-EXCHANGE GAME

---

##### Summary

In this second experiment we extend the gift-exchange game setting to include a new treatment where subjects are paired with the same partner for the whole game. It has been observed that trust-based contracts could be more efficient than incentive contracts when only one mechanism is available, whereas incentive contracts performed better than trust-based contracts when principals can choose either of the two contracts. However, these seemingly contradictory observations have been made in the context of stranger settings. Since most organizational processes involve repeated interactions in the long run between the same agents, the partner setting may provide a more accurate account of the respective performance of these two types of contractual arrangements. We do not observe that the incentive contract strictly dominates the trust contract in the stranger setting, while the trust contract strictly dominates the incentive contract in the partner setting.

## 1. Introduction

In the experiment presented in this chapter we investigate how the repetition of interactions between the same players influences contractual choices and cooperation levels over time. If some players actually integrate their long-term interests in their choice to cooperate or not, repeating the game should not only not lead to a general convergence to the stage Nash equilibrium but also generate some level of self-sustained cooperation. Accordingly, interaction mechanisms that do not allow for the provision of incentives may not necessarily be overtaken by those which do when the choice between the two is left to the players. If this would be actually the case, then it would be possible to relax - at least partly - the implicit assumption shared by most incentive-based theories that agents only take into account their short-term payoffs when they choose their effort or contribution levels. Relaxing this assumption could already provide a first step to a more integrated view of collective production processes.

We compare the experimental results of two gift-exchange game settings where principals can choose between two types of contract. In all contracts, principals must specify a wage and a minimum effort level. In trust contracts (TC) the minimum effort level is not binding for the agent who can then choose any actual level of effort without bearing any additional cost. In incentive contracts (IC) the principal can also choose a fine to be paid by the agent with a one-third probability if the minimum effort level is not reached. Establishing an incentive contract bears additional costs for the principal.

The gift-exchange game setting, which has been introduced in Chapter 2 (Section 4.1) is a good test bench for theories based on equity, fairness or reciprocity concerns (see Chapter 1, Sections 6.2 and 6.3). It has already been evidenced in many experimental investigations based on TCs that the agent's dominant strategy of choosing the lowest possible effort is by far not the dominant choice made by agents, and that a strong positive correlation between offered wage and chosen effort level can be observed. This contrasts with the IC case, where agents in experiments typically follow the dominant strategy of choosing the lowest possible effort<sup>47</sup>. These observations underline the complex interaction pattern between institutional designs and players' preferences, which appear to be distributed along a continuum between

---

<sup>47</sup> By "lowest possible effort" we mean here either the effort level specified in incentive-compatible contracts or the null effort level when the contract is not incentive compatible.

selfish, myopic, and purely monetary-driven utility functions at one end and altruistic, long-sighted, non-monetary driven utility functions at the other end.

To our knowledge, most experiments on gift-exchange games investigated a single interaction mechanism. This keeps the game relatively straightforward and thus allows to focus on the behavioral drivers at stake within a specific institutional arrangement. This however prevents performance comparisons between alternative institutional arrangements in situations where mechanisms are endogenously chosen by the players' themselves and not exogenously set by the experimenter. Furthermore, even a comparison across experimental sessions with exogenously given institutional mechanisms is hardly feasible because of the diversity of experimental outcomes considering the same mechanism and given the effects of the game structure on players perceptions and choices. Thus, the question of how trust-based contracts compare to incentive based contracts when chosen endogenously is still an open question. Current evidence suggests that trust based contracting is less beneficial for agents than incentive based contracts although higher wages are reciprocated with higher effort.

A few recent studies depart from this "single interaction mechanism" and implement alternative mechanism choices. Keser and Montmarquette (2004) introduced an experimental setting where players can choose between a private and a team remuneration in an effort game. Though it is a subgame perfect equilibrium strategy to choose private remuneration, team remuneration is chosen by both players in 45% of cases. Moreover, 85% of the players who were in the team remuneration mode one round chose that mode again in the next round and 52% of those who were in the private remuneration mode opted for the team remuneration mode subsequently. Fehr et al. (2007) investigate the gift-exchange game in a two-step setting. In the first step, principals chose the institutional arrangement, that is either TC or IC. Then, in the second step, the 'traditional' gift exchange game takes place. In their experiments, players play for several rounds but in each round agents and principals are matched randomly. They observed that IC tends to become the preferred arrangement over TC.

Our experiments follow the approach of Fehr et al. (2007) in that we use the same two-step procedure. We extend this work in that we investigate the same setting where principals and agents are matched randomly each round (in the following stranger treatment, ST), and compare it to a setting where the same principal-agent pair plays the game over all rounds (henceforth partner treatment, PT).



Our results are twofold. Firstly, observations on the stranger treatment deviate from Fehr et al. (2007) as we do not observe that IC strictly dominates TC. Instead, both arrangements are used by about half of the population throughout the rounds. Secondly and more importantly, TC strictly dominates IC in the partner treatment. Overall we find that principals' and agents' pay-offs are higher in TC than in IC in both treatments, wages and effort levels are higher in TC than in IC in the partner treatment, and all values are higher in the partner treatment than in the stranger treatment in both TC and IC.

Two preliminary conclusions can be drawn from these results. First, the efficiency of contract design is not clear cut *ex ante* and is mainly determined by agent-specific behavioral frames. It has not been observed here that the trust contract was outperformed by the incentive contract. Second, repeated interactions are not necessarily a source of learning towards some predefined equilibrium and can effectively support the *emergence of contract designs that economize on coercive mechanisms*.

In order to identify the underlying individual behavioral patterns that could explain the observed distribution of contracts we investigated for the partner and the stranger treatment how the principal's wage offer and proposed effort level affect the agent's realized effort, depending on the arrangement chosen by the principal. Presented results are obtained through the estimation of a linear regression model but we also considered non-linear and non-parametric models which yield essentially the same interpretation and are omitted here for simplicity. The differences between stranger and partner treatment are especially stark. In the stranger treatment, in both contract types, there is no systematic effect of proposed effort and wage on reciprocated effort. In the partner treatment both proposed effort and wage have a positive impact on reciprocated effort under the trust contract. Under the incentive contract the wage offered remains influential but the proposed effort becomes irrelevant for the agent's effort choice. Thus, the principal's choices in the gift-exchange game have a stronger influence on the agent's effort in the partner treatment than in the stranger treatment, and impact in the partner treatment further increases when a trust based contract is chosen instead of an incentive based contract.

These results on the individual level fit very well with the main patterns observed on the population level. In the stranger setting no arrangement is preferred because under neither arrangement the principal's choices effectively impact the agent's effort. The result is low benefit for principals and agents. In the partner treatment the trust contract dominates over

the incentive contract because under this arrangement agents respond to both, offered wage as well as proposed effort. Accordingly, gains are highest in the partner treatment under trust contracts.

## 2. The experiment

### 2.1. *The principal-agent games*

In all treatments, subjects played the same two-step principal-agent game. In the first step, the principal has to choose between the Incentive Contract (IC) and the Trust Contract (TC). In the second step, a gift-exchange game is played which differs by contract choice.

If TC is chosen the principal must specify a wage  $w$  and a desired effort level  $e^*$ . The agent then decides whether to accept or not the proposed contract. In the latter case the round is over and both agent's and principal's payoffs are null. In the former case the agent chooses an effort level  $e$ . The agent bears a non linear effort cost  $c(e)$ , increasing and with an increasing derivative in  $e$  (see Table 4.1). The cost and payoff functions are common knowledge.

The principal's expected payoff  $\Pi_p$  is given by:

$$\Pi_p = 10.e - w$$

The agent's expected payoff  $\Pi_a$  is given by:

$$\Pi_a = w - c(e)$$

If IC is chosen the principal must specify a wage  $w$ , a desired effort level  $e^*$  and a fine  $f$  to be paid by the agent with a probability  $p$  of 1/3 if the desired effort level is not reached. The agent then decides whether or not to accept the proposed contract. If the agent does not accept the contract the round is over and both agent's and principal's payoffs are null. Otherwise the agent chooses an effort level  $e$ . The principal bears an additional, fixed cost  $k$  to finance the control mechanism. The agent's effort cost function is the same as in the TC. Again, the cost and payoff functions are common knowledge.

The principal's expected payoff  $\Pi_p$  is given by:

$$\Pi_p = 10.e - w - k \text{ if } e \geq e^*$$

$$\Pi_p = 10.e - w - k + pf \text{ if } e < e^*$$

The agent's expected payoff  $\Pi_a$  is given by:

$$\Pi_a = w - c(e) \text{ if } e \geq e^*$$

$$\Pi_a = w - c(e) - pf \text{ if } e < e^*$$

Table 4.1: *Effort cost function for the agents.*

<b>e</b>	1	2	3	4	5	6	7	8	9	10
<b>c(e)</b>	0	1	2	4	6	8	10	13	16	20

## 2.2. *Game-theoretical solution*

If both players are assumed to maximize their monetary payoffs in each round and to believe that the other player will do so, the outcome of the stage game is straightforward. In the TC subgame the agent has a dominant strategy in choosing the minimum effort level ( $e = 1$ ) for any positive wage ( $w > 0$ ) and otherwise either choose  $e = 1$  or refuse the contract, each with a probability of  $1/2$ , so that the principal is left with the only option of choosing the null wage to secure the maximum possible payoff of 10. In the IC subgame the selfish agent will refuse the contract if  $w < c(e^*)$  and  $w < pf$ . The agent accepts the contract and chooses the minimum effort required  $e^*$  if  $c(e^*) < pf$  holds, else the agent chooses the minimum effort level  $e = 1$ . Since the highest available value for  $f$  is 13 the highest effort level that an incentive compatible contract can elicit is thus  $e^* = 4$ . It is thus a strictly dominant strategy for the principal to choose a wage  $w^* = 5$ , a required effort level  $e^* = 4$  and the maximum fine  $f = 13$ , a contract to which the agent will reply by choosing the required effort level of  $e^* = 4$ . The latter yielding a payoff of 25 to the principal, the IT will be the only contract chosen for a payoff of (25,1) to the principal and the agent respectively.

The social optimum, which requires that the principal chooses the TC and the agent accepts and chooses the maximum effort level of 10 is thus precluded, as well as any intermediate

outcome where the agent a positive effort level in a TC or an effort level higher than the required one in IC.

Furthermore, there is theoretically no difference between the Partner and the Stranger treatments with the above assumption on selfish and myopic players, since in both cases the timeline is bounded. However, repeated interactions with the same players allow for reciprocity effects to be at play and have been experimentally related to some positive level of voluntary cooperation, either triggered by reciprocity on observed behavior ("reaction-function" reciprocity) or by believes on others reciprocal motives. For instance, results from the Bonus treatment<sup>48</sup> in Fehr et al. (2007) clearly indicate that adding one step in the decision process could trigger higher effort levels. Accordingly, the Partner treatment should offer more opportunities to both principals and agents to signal their willingness to cooperate and thus favor the TC.

### 2.3. *Experimental settings*

The experiment was carried out at the Laboratoire d'Economie Expérimentale de Strasbourg (LEES) using our in-house designed data-processing software. A total of 80 voluntary subjects took part in the experiment after being randomly selected through ORSEE among 1200 students from various programs.

None of them had previously confronted gift-exchange game experiments (inexperienced subjects). Written instructions were distributed and read aloud to the subjects before they performed a pre-experimental test to check proper understanding. The session began as soon as they all had correctly answered every question. No communication was allowed between subjects as long as the experiment was running.

We applied two treatments, namely a Stranger (S) treatment and Partner (P) treatment. In both treatments we formed randomly two groups of ten subjects who played as principals and agents respectively. In both treatments subjects played for ten rounds. In the Stranger treatment we formed new principal-agent pairs each round, each player being paired once with each of the ten players of the other group. In the Partner treatment we formed ten principal-agent pairs, each player being paired with the same partner all ten rounds. Each subject played the same role (principal or agent) all along the game and participated in only one session and thus in only one of the two treatments, so that data are independent from one

---

<sup>48</sup> In the Bonus treatment the principal could specify an additional amount, the bonus, to be possibly transferred to the agent.

treatment to the other and pair data in the Partner treatment are independent from one pair to any other. Table 4.2 summarizes the experimental treatments.

*Table 4.2: Experimental settings.*

Treatment	Sessions	Number of subjects	Number of pairs	Data points
Stranger	2	40	200	400
Partner	2	40	20	400
Total	4	80	-	800

At the end of each round both players would be displayed a result screen where the choices they made and their own payoff for the round appeared. At any moment during the experiment players could access their own history table with the above mentioned information for each previous round and their own cumulated payoff.

### 3. Results

Results are organized in four subsections. The first subsection, 3.1, compares the overall use of contract choices in the Partner and the Stranger Treatment. Sections 3.2 and 3.3 focus on the aggregate level. Section 3.2 compares the outcomes, or contract efficiency, by treatment, and Section 3.3 compares the outcomes by contract. In Section 3.4 we investigate the agents' response to the principals' choices and offers on an individual level.

#### *3.1. Contract choices : Partner Treatment (PT) vs Stranger Treatment (ST)*

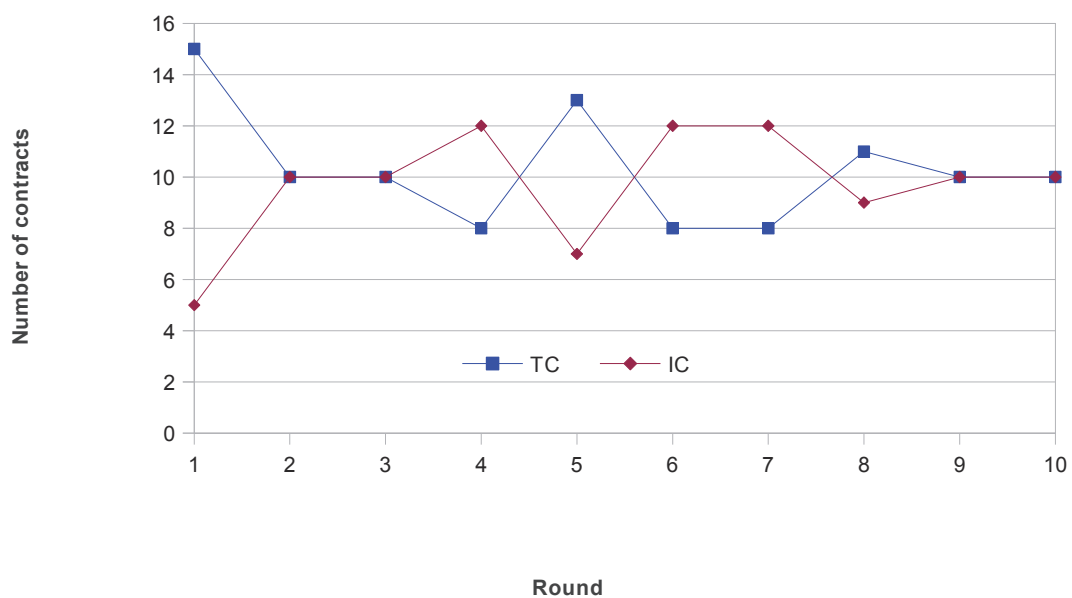
We compare here the relative use of TC and IC in each setting, overall as well as over rounds. Furthermore, we investigate how principals contract choices switched over rounds, which is a clear indicator of relative performance and learning.

**Table 4.3:** *Contracts shares and average outcomes.*

Treatment	Contract	Share	Wage	Effort	Principal payoff	Agent payoff
Stranger	TC	52%	15.2	1.9	3.2	13
	IC	51%	16.4	2.3	-1.3	11
Partner	TC	72%	31.2	5.3	12.7	22
	IC	28%	26.4	3.7	0.5	17.9

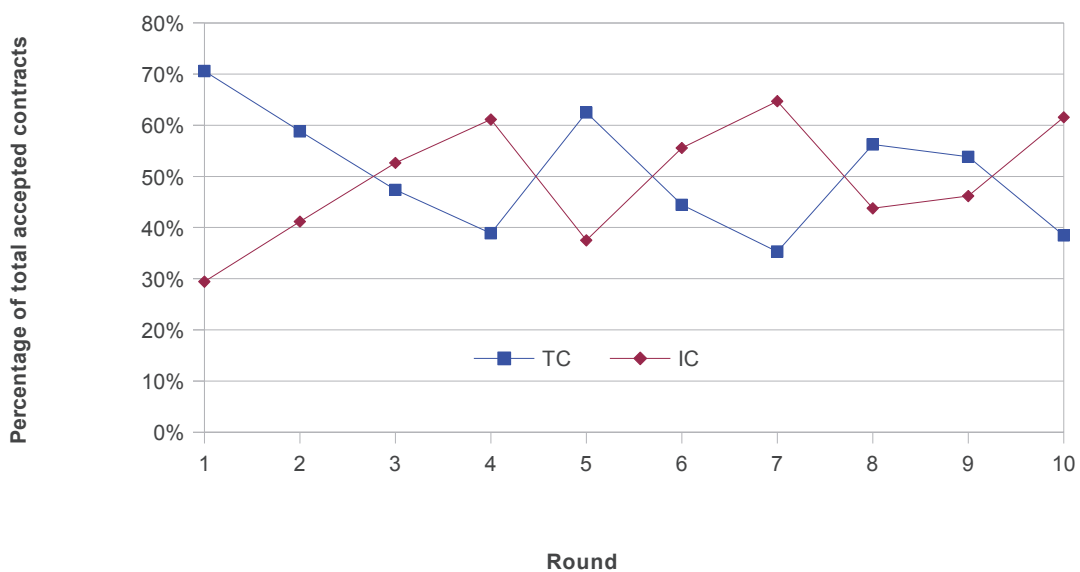
*Shares of IC vs TC*

Table 4.3 displays the share of contract types by treatment jointly with the average decisions and outcomes by treatment and contract type. Contrasting with the results of Fehr et al. (2007) we do not find that IC dominates TC in the Stranger Treatment (see Figures 4.1A and 4.1B). Instead, both contract types are used at about the same frequency. Out of 200 offered contracts, 103 (52%) were TC, and 51% of accepted contracts (51%) were TC (see Table 4.3). We observed that TC was preferred to IC in the first round (70%), and that both contracts held a closely even share of principals' choices in the following rounds. In rounds 2

**Figure 4.1A:** *Stranger Treatment - TC and IC contracts chosen by the principals, by round.*

and 3 as well as in the two final rounds the principals' choices led to a 50% divide between the two contracts. TC have been slightly more refused by agents than IC: 83 TC (81% of the proposed TC) and 81 IC (84% of the proposed IC) have been accepted. Two third of contract rejections have occurred in the last five rounds.

Figure 4.1B: *Stranger Treatment - Share of accepted TC and IC in percentage, by round.*



In the Partner Treatment TC strongly dominated IC. TC was chosen in 143 cases (72% of all 200 cases) and 143 accepted contracts were TC (70% of the 159 accepted contracts). Figures 4.2A and 4.2B show that TC was chosen in 90% of cases in round one, that more principals chose TC in all rounds, and that TC was chosen in at least 70% of cases from round 5 on. The lowest share of TC was 56% of accepted contracts in round 2 and 55% of proposed contracts in round 4. Here again the percentage of accepted to proposed IC is slightly higher, with 78% of TC accepted by the agents against 84% for IC. From round 7 on though, the percentage of accepted TC is equal to or higher than the percentage of accepted IC.

Figure 4.2A: Partner Treatment - TC and IC contracts chosen by the principals, by round.

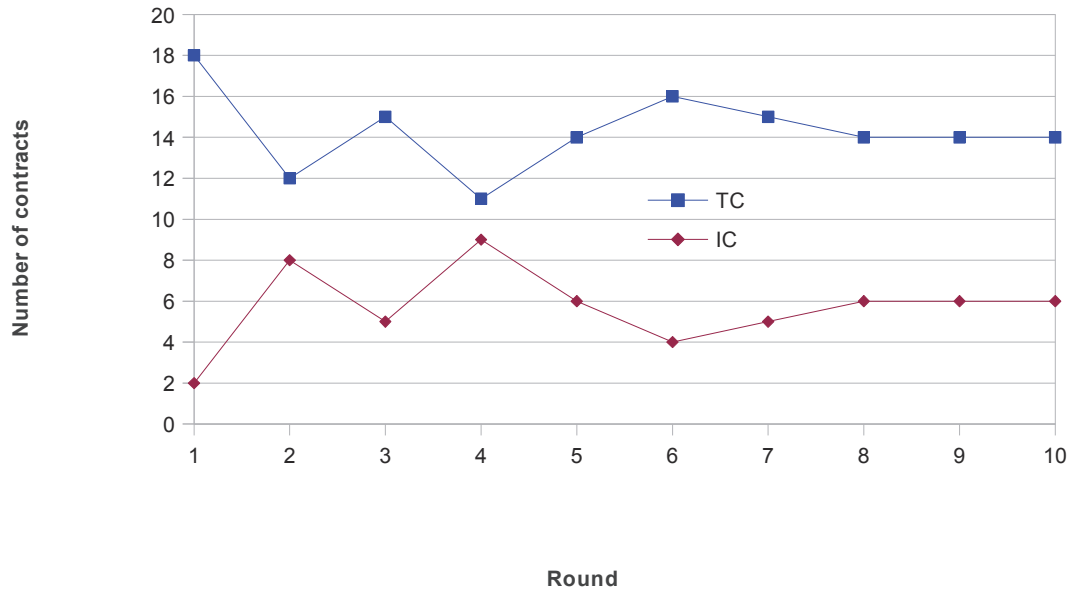
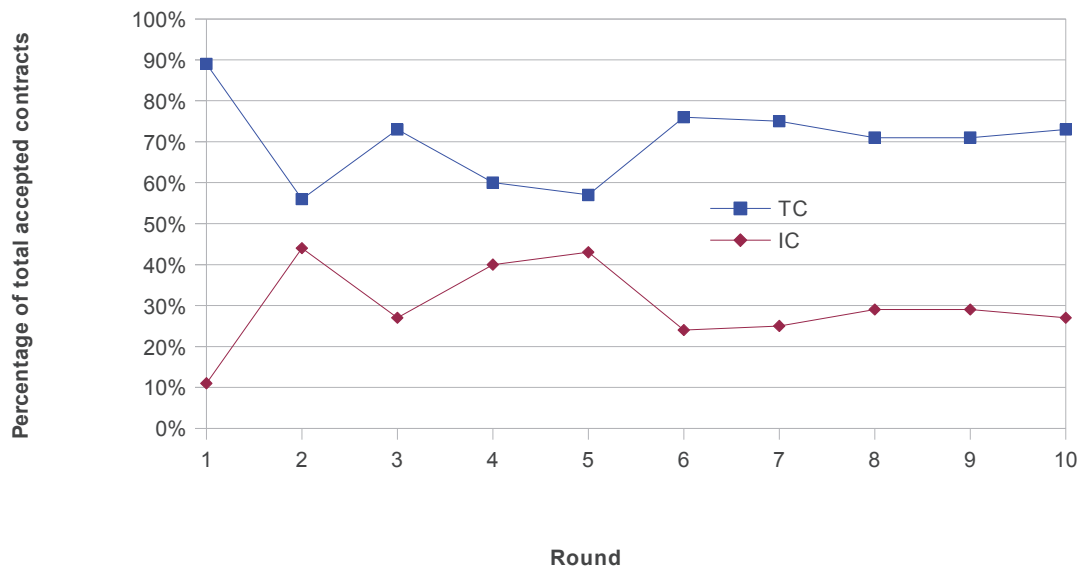


Figure 4.2B: Partner Treatment - Share of accepted TC and IC in percentage, by round.

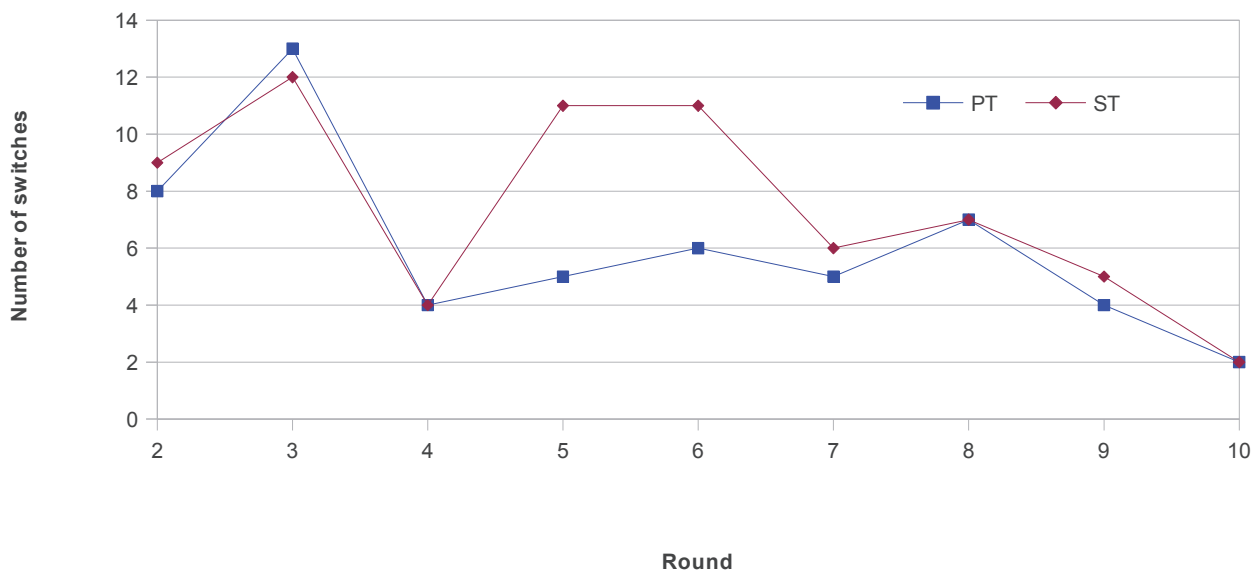




*Contract switches: the relative stability of contract choices*

Since the repeated game setting provides the players with an opportunity to learn in both treatments as well as a signaling tool in the Partner setting, it is worth examining the pattern of contract switches to investigate whether principals did stick to their first round choices or did experiment with the other contract.

Figure 4.3: *Number of round-to-round contract switches*



In each treatment a total of 180 choices were made where the principal could choose between keeping the same contract and switching to the alternative contract. In the ST 67 switching choices were observed, as compared to 54 in the PT. The evolution of the number of switching choices (see Figure 4.3) shows a similar pattern in both treatments for the four first and four last rounds, with a higher proportion of switches in the first rounds and a clear decrease in the last rounds. In the ST, however, we observed a strong increase in the number of switches in rounds 5 and 6, amounting to 11 out of 20 choices for each of these two rounds or twice as much as in the PT. In the 5th round, 8 of the 11 switches were from IC to TC, 6 of which were reversed in the next round where a total of 8 switches were from TC to IC. Principals in the Stranger treatment have kept switching contracts and exploring long after the first rounds, and have reversed their choices in a larger extent than principals in the Partner treatment later on in the game. The observed indecisiveness in ST is also comforted by the share of the least chosen contract for each principal, which is 40% or 50% in 6 cases in ST as compared to only 3 cases in PT. Similarly, only one principal did stick to his first

round choice (TC) in all rounds in ST as compared to 5 in PT. We consider this instability of contract choices in the Stranger treatment to give strong evidence that *none of the two contract types holds an absolute competitive advantage on the other* whatever the individual rationale for choosing one rather than the other.

Over the two treatments, 6 principals thus chose the same contract, TC, in all rounds (1 in ST and 5 in PT). Among the 7 principals who chose a different contract only once, 5 chose TC (2 out of 4 in ST and 3 in PT), which was also chosen by 6 principals among the 9 ones who chose a different contract twice (3 out of 5 in ST and 3 out of 4 in PT). Furthermore, 72% (13 out of 18) principals who chose TC in the first round in PT did mostly stick to their first-round choice, choosing TC in 7 or more rounds throughout the whole game. These observations further underline the relative advantage of TC over IC, particularly sharp in a partner setting, since most principals who exhibited a sustained preference for one of the contract chose TC. By contrast, no principal chose IC in all rounds in any of the treatments.

### 3.2. *Contract efficiency: Partner Treatment (PT) vs Stranger Treatment (ST)*

Table 4.4 displays the comparative values of offered wages, chosen effort level, and principals' and agents' payoffs, averaged over all rounds and all concerned players<sup>49</sup> for the two treatments. Average offered wage was much higher in the PT (29.8) than in the ST (15.7). At first look, agents seem to have responded to the higher wages in PT with an average effort level of 4.6 against 2.1 in ST. The minimum possible effort level being 1, the difference between the two treatments in terms of effort level is even stronger than the difference in terms of average offered wages. We indeed observed a correlation coefficient of 0.73 between wages and chosen effort level in PT, as compared to 0.21 in ST. The outcome was a strong difference between the two treatments both in terms of principals' average payoff, which was 0.9 in the ST against 9.2 in the PT and of agents' average payoff, which was 12 in the ST and 20.9 in the PT.

The PT had a positive effect on payoffs values for both players and for both types of contracts, the average payoff being between 1.8 and 9.5 points higher in PT than in ST, depending on player type and chosen contract.

As expected in the light of the previous observations on contract shares, the PT was more favorable to the TC also in terms of its relative efficiency in comparison to the IC. The

<sup>49</sup> The analysis of the average effort does not include the incomplete cases where the proposed contract was refused by the agent. Unless otherwise specified, all the differences in average values presented below are significant at the 0.05 confidence interval (Mann-Whitney rank-sum test).

difference between PT and ST in terms of average wages and effort level is greater for the TC than for the IC, in such a way that the ranking of the two contracts in these terms is reversed from PT to ST.

*Table 4.4: Contracts shares and average outcomes by treatment and by contract.*

Contract	Wage			Effort			Principal payoff			Agent payoff		
	ST	PT	All	ST	PT	All	ST	PT	All	ST	PT	All
TC	15.2	31.2	24.5	1.9	5.3	3.8	3.2	12.7	8.6	13	22	18
IC	16.4	26.4	20.1	2.3	3.7	2.8	-1.3	0.5	-0.64	11	17.9	13.6
Together	15.7	29.8	22.75	2.1	4.8	3.4	0.94	9.2	5.07	12	20.9	16.45

### 3.3. *Contract efficiency: Incentive Contract vs Trust Contract*

We now turn to a closer examination of the relative performance of the two types of contracts, whether IC and TC were differently accepted or not, and whether averages and tendencies of offered wages, effort levels, principals pay-offs, agents pay-offs, and social output were similar or not.

We noted that IC were slightly more accepted than TC in both treatments. 84% of the proposed IC were accepted in each treatment against 81% and 78% of the proposed TC, in the ST and in the PT respectively.

As shown in Table 4.4, average offered wages in IC and TC were about the same level within each treatment, still the two types of contract performed in the reversed order from one treatment to the other, : in ST wages proposed with IC were around 7% higher on average than proposed wages in TC<sup>50</sup> whereas in PT wages proposed with TC were around 18% higher than in IC. The average wage offered with TC over both treatments was 24.5 against 19.5 for IC. Wages offered with both contracts decreased over time in ST, from an average wage of 32.9 for TC in the first round down to 5.6 in the last round and from 28 down to 13 for IC. An even steeper decrease was observed in PT for the IC, from an average first round wage of 48 down to 17.7 in the last round, however the TC average wage in PT exhibit a rather moderate decrease, from 35.4 to 29.4.

<sup>50</sup> The difference here is not significant.

Effort levels followed the same pattern as wages, the average chosen effort level being higher for IC than for TC in the ST and vice versa in the PT (see Table 4.3). Over the two treatments however the average observed effort level was 3.9 in TC against 2.8 in IC. The higher offered wages in TC seem thus to have successfully triggered higher effort levels. It has moreover been quite effective in doing so since the observed increase of 26% of the average wage was associated with an increase of 38% of the average effort level. The correlation coefficient between wages and effort level is indeed 0.72 for the TC over the two treatments, against 0.54 for the IC. The difference between the two contracts in terms of effort levels is quite stable over time, the average effort level being higher in TC than in IC in all rounds excepted round 6 where the average effort level is the same for the two contracts.

The effort levels were rather stable over time, the only significant decreasing trend (Spearman rank correlation between round and average effort level  $R = -0.615$ ,  $p = 0.058$ ) being observed for the average effort levels chosen in IC in the ST.

In both treatments, the average payoff was thus higher in TC for both types of players. The average payoff for principals in TC was 8.6 against -0.6 in IC, agents obtain an average of 18 in TC and 13.6 in IC. As already mentioned, the PT was relatively more favorable to TC. However, the collective cost of IC, where principals had to pay  $k = 10$  for the control device and agents had to pay on average one third of the fine in case they chose an effort level lower than the requested one, made the TC more efficient than the IC in both treatments in terms of payoffs: in the ST the average payoff for principals was 3.2 in TC against -1.3 in IC, when the average agent's payoff was 12.9 against 11. The *social outcome* was thus higher on average in TC than in IC, in each treatment.

Finally, principals were better off on average under TC in the ST than they were under IC in the PT. This provides some additional evidence that the mechanisms supporting the emergence of cooperation through repetition may not necessarily add up with the effects of incentives, since the latter even seem to be counteracting the benefits of repetition between the same pairs.

### 3.4. *Reciprocity: the determinants of effort levels*

We will now try to get a clearer picture of the interaction between the possible determinants of effort level - wage, required effort level, treatment and contract. The analysis of the comparative responsiveness of agents to offered wages and required effort level by treatment

and by contract should indeed provide some explanation to the contract shares observed at the aggregate level.

We first ran a linear OLS regression for the simple following model:

$$e_{it} = \alpha + \beta \cdot w_{jt} + \gamma \cdot e^*_{jt} + \delta \cdot d_{IC} + \varepsilon_{it}$$

where  $i$  and  $j$  are respectively the agent and the principal paired at round  $t$  and  $d_{IC}$  is a dummy for the chosen contract, with  $d_{IC} = 1$  if the Incentive Contract is chosen and  $d_{IC} = 0$  otherwise.

The estimates of this model (see upper part of [Table 4.5](#)) already point to a striking difference between the two treatments, showing a strong and significant positive effect of the offered wage and a smaller, weaker but also positive and significant effect of the required effort level in the Partner treatment whereas none of the two variables exhibit a significant effect in the Stranger treatments.

In order to grasp the interaction effect of the chosen contract with the two other variables we then ran a slightly refined model where we combined the contract dummy with the offered wage and the required effort level:

$$e_{it} = \alpha + \beta_1 \cdot w_{jt} \cdot d_{TC} + \beta_2 \cdot w_{jt} \cdot d_{IC} + \gamma_1 \cdot e^*_{jt} \cdot d_{TC} + \gamma_2 \cdot e^*_{jt} \cdot d_{IC} + \varepsilon_{it}$$

where  $d_{TC}$  is an additional dummy for the chosen contract, with  $d_{TC} = 1$  if the Trust Contract is chosen and  $d_{TC} = 0$  otherwise.

As can be seen in the coefficients displayed in the lower part of [Table 4.5](#), the effort level chosen by the agents was significantly positively correlated with the offered wage in both contracts in the Partner Treatment. The required effort level also seems to have a significant effect but only in the TC. We can also observe that the chosen contract *per se* did not significantly affect the effort level, as shown by the contract dummy in the simple model, and that wage and required effort level were determinant in the efficiency of the TC. It appears thus that intentionality models of reciprocity (see e.g. Rabin, 1993, or Falk and Fischbacher, 2006, presented in section 6.3 of Chapter 1), which would have attributed an effect to the choice of a punishment-free contract, may not add much in the explanation of agents choices in our case.

Here also, as with the simple model, none of the explanatory variables is associated with a significant coefficient in the Stranger treatment. This result partly stems from the large number of agents who chose the minimum effort level in this treatment, as well as from the rather low number of incentive-compatible contracts among the chosen IC. As in Fehr et al.

(2007) we observed a larger share of these in the last round (50% - four out of a total of eight IC) than in the first round (20% - one in a total of five IC), though overall only 24% (19 out of 81) of the accepted IC used such values of  $w$ ,  $e^*$  and  $f$  that the best response for the agent in terms of monetary payoffs was to choose the required effort level. This explains why we do not find a systematic effect of wages and required effort on chosen effort in the Stranger Treatment, where IC was chosen more often than in the Partner Treatment.

*Table 4.5: Estimates of the linear model OLS regression (Std error).*

	Stranger		Partner <sup>51</sup>	
Intercept	1.35 *** (0.34)	1.44 *** (0.32)	0.67 (0.80)	-0.02 (0.67)
$w$	0.003 (0.009)	-	0.11 *** (0.01)	-
$e^*$	0.1 (0.06)	-	0.17* (0.09)	-
$d_{IC}$	0.27 (0.30)	-	-0.7 (0.45)	-
$w_{jt} \cdot d_{TC}$	-	-0.0006 (0.009)	-	0.09 *** (0.02)
$w_{jt} \cdot d_{IC}$	-	0.03 (0.03)	-	0.13 *** (0.02)
$e^*_{jt} \cdot d_{TC}$	-	0.11 (0.07)	-	0.23 ** (0.10)
$e^*_{jt} \cdot d_{IC}$	-	0.04 (0.09)	-	0.02 (0.12)
Adjusted R <sup>2</sup>	0.02	0.02	0.54	0.38

\*\*\*  $\Pr(>|t|) < 0.001$     \*\*  $\Pr(>|t|) < 0.05$     \*  $\Pr(>|t|) < 0.1$

<sup>51</sup> As could be expected in the Partner treatment we found a small but significant correlation of residuals over time. However, applying a fixed effects models over pairs did not significantly change our results. The random effect model being more efficient in our setting, and the Hausman test ( $p=0.2759$ ) suggesting that both models are unbiased, we relied on the estimates of the random effect model. Though similar issues might in principle arise with the Stranger data, the results of the general model did not find any significant

The addition of a round dummy did not significantly change the results from the static model for the Partner Treatment, confirming the results from the above non-parametric analysis (section 3.3) where no trend could be identified. Taken together, these results support a strong effect of the repetition of interactions between the same pairs of subjects, through a sustained responsiveness of agents to both  $w$  and  $e^*$  when no incentives are set.

#### 4. Discussion

Our results clearly indicate that the TC is more favorable to cooperative outcomes than the IC. However, as implemented here the IC contract is cognitively more complex, since it requires principals to set correctly the required effort level, the fine and the wage parameters in order to achieve incentive compatibility. The large number of non incentive-compatible IC and the increasing share of incentive compatible ones shows that this has not been an easier task for our subjects than for Fehr et al. (2007) subjects<sup>52</sup>. One could also point that the IC could not by design elicit higher effort than  $e = 4$  and that it was associated with a fixed cost. We also observed that most of the incentive-compatible IC were successful in determining the agent to choose the specified effort level, but that in only one case a higher effort was provided.

In Fehr et al. (2007) the authors claim that "the principals converge toward the most efficient contract in the set of available contracts", which is at least partly backed by their data but is contradicted by the results of our Stranger treatment, in which we reproduced their Trust-Incentive treatment as a benchmark for the Partner treatment. We observed that TC and IC hold an equal share of principals' choices until and including the last round of the experiment. We agree with them that "this observation is important because the "efficiency principle" provides the basis for much of modern contract theory" but in both treatments the TC happened to be on average more efficient than the IC. Though it is outside of the scope of this chapter, it could be interesting to investigate how the inequity aversion model they used would help explain and interpret these differences, since they actually found a very close match between their experimental results and predictions based on a fixed set of parameter values applied to their model.

---

effect of the studied variables so that further analysis was irrelevant.

<sup>52</sup> "Over time, most principals learned to make the contract incentive compatible, but this was not a trivial task. After all, no principal observed what the other principals did, so everybody had to figure it out on their own. Not all principals managed to do so within the ten rounds of the experiment. "

The only differences between their settings and ours were that (i) they used the connoted terms "employer" and "employee" in the instructions where we used the neutral expressions "player A" and "player B", (ii) they relied on a manual procedure with paper and pen without isolating players<sup>53</sup> while we used computer terminals and blinders, (iii) their subjects were students of "the natural sciences, engineering, law, political science and mathematics from the University of Munich and the Technical University of Munich" while our came from more diverse fields of the University of Strasbourg and (iv) they explicitly required that the offered wage cover the effort cost, which we did not. Since the incentive compatibility of the chosen IC is crucial to their performance, any procedural factor that may favor the correct calculation could partly explain the difference between our results in the Stranger Treatment and theirs.

Finally, we can also notice that the Bonus treatment in Fehr et al. (2007), where principals choose between IC and a Bonus Contract (BC), yields very similar results to our Partner treatment. In the BC, principals must specify an amount that they intend to pay if the required effort level is matched. The specified amount is only indicative and principals can choose to pay any amount, or nothing, after being notified of the agents' choices. In this setting, thus, the principals have also the opportunity to respond to the agents' effort level choices. In both the Bonus treatment and our Partner treatment it is obviously less costly for principals to trust the agents.

## 5. Conclusion

We have let two alternative mechanisms differentiated along the lines of the two conceptions of cooperation compete in a small population of principals and agents. When the interactions were not repeated between the same pairs of subjects, the Incentive Contract and the Trust Contract shared equally the principals' choices from the second round on. When the same players were matched for all rounds, the Trust Contract was chosen by a majority of principals and in more than 70% of cases in the second half of the timespan.

These results support the conception of cooperation as possibly emerging from non-binding settings where agents must rely on reciprocal trust to produce collective outcomes. Importantly, this form of cooperation is enhanced by the repetition of interactions between the same agents, which is usually the case in organizational processes. It does not preclude

---

<sup>53</sup> They describe in the Appendix how principals who had chosen the Trust Contract were also throwing a dice not to let the others know their contract choice.



that incentives can indeed trigger cooperation efforts but these should not be expected to go beyond incentive compatibility.

We also found some evidence that a large part of the agents cooperative choice can be explained by the perspective of lasting payoffs when subjects are paired for the whole timespan, so that the minimal form of trust - trusting that the other players have also identified the long term benefits they can expect from cooperation - can be enough to sustain spontaneous cooperation.

This furthermore raises the issue of the availability of alternative designs. If cooperation is influenced by existing contract designs, then the dominant form of contract design will tend to influence the extent to which emerging or elicited cooperation is perceived as the normal form of interaction.

---

## CHAPTER V

### DISENTANGLING DISTRIBUTIONAL MOTIVES

---

#### Summary<sup>54</sup>

According to our starting definition, cooperation involves the achievement of an outcome that cannot be achieved alone. The so-called 'non-cooperative' games that have been used up to this point to study the extent to which cooperation could emerge in an incentive-free environment are monetary-only replicates of the production processes that they are intended to sketch out, so that it is not possible to clearly distinguish between the distributional implications of the various possible outcomes and the efficiency achieved through cooperation. In this chapter we present the result of a distribution experiment where players must choose between a maximin, equity-dominant solution and a Hicks optimal, efficiency-dominant distribution. Three different information conditions are used. Inequity aversion have no observable effect on players' choices in the certain information condition, while the two other conditions yielded more contrasted outcomes.

---

<sup>54</sup> The experiment presented in this chapter was prepared in collaboration with Herraide Igersheim.

## 1. Introduction

Numerous experiments have shown that other-regarding preferences do influence players' choices. However, in the experiments exploring cooperation presented in the previous chapters the collective output is simultaneously produced and distributed, so that cooperative choices enhancing collective production and distributional preferences are entangled and it is often difficult to clearly identify individual motivations pertaining to one or the other.

In pure distribution games no strategic interactions are at stake and the efficient outcome can be differentiated from the cooperative outcome. Cooperation here is not identified *ex ante* since it is possible to consider either choice as a form of cooperation, towards the efficiency-dominant outcome maximizing the total group revenue in one case, or towards the equity-dominant outcome reducing payoff discrepancy within the group in the other case. Since both, or neither of these principles may be active in non-cooperative game settings it is necessary to study them separately in order to be able to tighten the attribution of cooperative choices to the relevant factors.

## 2. Related literature

The distribution games presented in Chapter 2 (Sections 4.4 and Section 5) are the closest attempts to carefully isolate and measure the weight of other-regarding preferences in experimental games.

### 2.1. *Ultimatum and dictator games*

The results of two-player dictator game experiments provide a first approximation of the willingness of subjects to reduce their own monetary pay-off in order to reduce the difference between their own pay-off and that of the other player. However, as previously mentioned in this thesis, transposing results from two-player settings into multiple-player settings is not straightforward and cautiousness requires to keep most parameters constant while exploring any specific dimension.

In the three-player ultimatum games presented in Güth and van Damme (1998) and Bereby-Meyer and Niederle (2005), where one of the players, the 'dummy', cannot voice his preference, the outcome is on average worse for the dummy than the outcome of the dictator

game is for the second player. Here again, though three players participate in the game, the outcome is the result of the interaction between two of them and of their consecutive choices.

## 2.2. *Efficiency, maximin and the two models of inequality aversion*

Engelmann and Strobel (2004), hereafter E&S, specifically address the relative predictive power of equity theories in three-player settings. They compare (i) maximin, (ii) efficiency (Hicks optimality), (iii) equity as defined by Bolton and Ockenfels (2000) in their theory of Equity, Reciprocity and Competition (ERC) and (iv) equity as defined by Fehr and Schmidt's (1999) theory of Fairness, Competition and Cooperation (FCC). In each of the games they used, players had to choose between three distributions determining the payoff to the three players in the group, two of these distributions being the solutions pointed to by one or more of the above theories.

The *maximin solution* simply involves choosing the distribution where the minimal payoff is the highest between the three, no matter the differences between each player's payoff and the payoff of any other player and no matter the total payoff for the group. The '*efficient solution*', in the sense of E&S, is defined according to Hicks optimality and is thus borne by the distribution in which the sum of the payoffs of the players is the highest, regardless of any other consideration. Both maximin and efficient solutions are based on the overall outcome for the group and assume thus that players' preferences are fully determined by considerations pertaining to the collective situation only. This assumption can be sustained in relation to one of the features of the experiment in E&S, namely that they compare only the choices made by the one player whose monetary payoff is not affected by the distribution, in order to keep selfish considerations out of the picture.

ERC and FCC are both based on a trade-off between selfish behavior - the player's own monetary payoff - and inequity aversion. Since the monetary payoff to the player whose choice is under scrutiny is kept constant through the three distributions, ERC and FCC solutions are determined by their respective treatments of inequity aversion. The *ERC solution* relies on the lowest difference between the player's payoff and the average payoff to the players in the group. The *FCC solution* in contrast is determined by the lowest absolute total difference between the player's own payoff and each of the other players' payoffs. In ERC, subjects' choices are thus again determined by the group outcome, since they are comparing their situation to a characteristic of the collective situation, whereas FCC assumes that subjects are choosing according to the one-to-one relationship between players' payoffs.

All subjects in E&S had to indicate their preferred distribution as 'Person 2', the choice of which would determine the outcome for the three players selected to be in the same group. Groups were formed and roles attributed after subjects had indicated their choices, so that subjects did not know if their own choices would be picked up nor did they know the identity of the other players in their group. The former uncertainty on their situation in the group was used in order to "generate three times the data" since all subjects played 'as if' they would be Person 2, but may have generated some bias as non-certain situations usually do.

The matrices were designed in order to isolate one or more of the solutions from the other one(s) and to implement some variations in the choosing player's situation in order to generate more detailed results. The overall results indicate that ERC and FCC solutions are chosen only by a very small portion of the players (between 3% and 7% depending on the matrix) when either of the two points to a different distribution from the one bearing the three other solutions. By contrast, each of the maximin and efficiency solutions gathers a substantial amount of choices, respectively 53% and 40%, even when the three alternative solutions all support a different distribution. Thus, in the condition where the payoff matrix has one distribution supported by the maximin solution alone, the latter is chosen by a majority of the subjects, seemingly making maximin considerations the strongest motive in this distribution game series. This condition requires that the choosing subject gets the highest payoff in all cases, so the results imply that a majority of subjects tend to be less sensitive to overall inequity when it positively affects the difference between their own payoff and the other players' payoffs while favoring the lowest difference among other players. In all conditions where the maximin and the efficiency solutions supported the same distribution, not supported by the equity solutions, that distribution was chosen by 60% to 77% of subjects. When efficiency and ERC, on the one side, and maximin and FCC, on the opposite side, support the same distributions, these two distributions yield similar shares of the subjects' choices, around 40% each, while the intermediate distribution still gets 20%. E&S then conclude that maximin and efficiency can account for most of the observed outcomes, FCC and ERC solutions doing fairly poorly in comparison.

However, keeping the monetary payoff equal across distributions for the choosing player may result in magnifying the actual importance of the studied alternative components of players' choices, or distort their actual relative roles when the player's own monetary payoff is also determined by the outcome. Though it was a deliberate choice intended to isolate the motives under study from purely individual monetary stakes, it missed a crucial part of the whole interaction between motives. Furthermore, some characteristics of the distributions

may be more or less salient, and thus exert more influence on choices, in some matrices, than originally designed, and subjects may also be more or less sensitive to absolute values. Finally, subjects were choosing 'as' Player 2 but did not know whether their choices would be determining the outcome, nor what role they would be playing.

**2.3. A renewed case for inequality aversion?**

In their answer to the previous paper, Bolton and Ockenfels (2006) introduced a majority rule voting game where three players must indicate their preferences between two distributions. In order to extend the study of the influence of equity on players' choices they also introduced a form of procedural equity in a second condition - 'equal opportunity mode' - where subjects faced role uncertainty, in addition to a 'straight mode' in which subjects knew their own roles before indicating their preferences. The 'equal opportunity mode' features an equivalent of the strategy method where players were asked to mark their preferred distribution conditional on the role that they would be randomly assigned afterwards. This condition allowed them to test the influence of procedural equity on preferences though it left open the question of the general preferences of subjects for one or another distribution.

*Table 5.1: Game payoffs* (reproduced from Bolton and Ockenfels, 2006).

Game	All	Game I	Game II	Game III
Distribution	A	B		
Player 1	13	19	27	27
Player 2	13	13	1	9
Player 3	13	13	17	9
Total	39	45	45	45

In all treatments, distribution A is an egalitarian distribution which allocates 13 to all three players whereas distribution B is an uneven distribution that yields a higher total payoff to the group. Table 5.1 above displays the payoffs associated with distribution B in each of the three different games used in the experiment. While the efficiency solution is always distribution B, distribution A bears the equity solution in all three games. The maximin

solution is also borne by distribution A in Game II and Game III, both distributions yielding the same minimum payoff in Game I.

We will first focus on the straight mode. Contrary to the settings used in E&S, the distribution chosen had an influence on the monetary payoffs to the players in most cases. In Game II and Game III, all players monetary payoffs are different from distribution A to distribution B, and since the preferences of all players determine the chosen distribution, and are thus reported in the results, the observations on Player 1 choice in Game I are also driven by preferences on the absolute value of the player's own payoff.

The results of Game I provide a sound test of the role of inequality aversion when the difference between monetary payoffs is to the disadvantage of the player. 48% of subjects whose payoffs were the same in the two distributions (Player 2 and Player 3) indeed chose distribution A though distribution B yielded lower payoffs to none of the players.

However, when the monetary payoff to the player varies according to the chosen distribution the shares of players' choices are substantially less balanced. Distribution B was preferred by 75% of the Player 1 subjects in Game I, 67% in Game II and 79% in Game III, and by 63% of Player 3 subjects in Game II. Similarly, distribution A was chosen by 88% of subjects in the three cases where it was more favorable to them than distribution A - Player 2 in Game II, Player 2 and Player 3 in Game III. A large majority of subjects thus chose the distribution in which their monetary payoffs were the highest, with marginal adjustments according to the absolute and relative values of the other players' payoffs in the case of Player 1. The highest variation between Player 1 subjects' choices, for instance, was observed from Game II to Game III, the latter featuring a no more equitable distribution B than the former in both of the above definitions, a higher minimum payoff, but also a more even situation in the sense that Players 2 and 3 get the same payoff instead of Player 3 getting almost all of these two players' shares. A similar outcome is observed with one of the matrices used in E&S, as shown in [Table 5.2](#). A majority of subjects, in the role of Player 2, chose the least equitable, least collectively profitable distribution C, which dominates the two others in terms of lowest payoff (maximin) and again provides the most leveled monetary payoffs to the two other players. Subjects may thus also be motivated by impartial equity concerns by which they favor the lowest difference between the other players' payoffs without reference to their own payoff.

Table 5.2: Payoffs and player 2 choices (Treatment R, Engelmann and Strobel, 2004).

Distribution	A	B	C
Player 1	11	8	5
Player 2	12	12	12
Player 3	2	3	4
Total	25	23	21
Share of choices	27 %	20 %	53 %

The 'equal opportunity mode' mostly affected the choice of Player 2 and Player 3 subjects in Game I. The share of these subjects choosing distribution A shrunk from 48% down to 17% when all three subjects in the group had an equal probability to be in each of the three roles. Keeping in mind that in this condition subjects would vote for their preferred distribution for each of the possible roles, the difference can only be attributed to the introduction of procedural equity and thus strongly supports the effect thereof. It also gives another confirmation that preferences can be substantially modified by the very procedure by which subjects express them, as already observed in Shafir and Tversky (1992) and in Frohlich and Oppenheimer (1998) in the case of prisoner's dilemma games.

### 3. A simple distribution game

#### 3.1. Refining the settings

In order to carry the elucidation of the role of distributional concerns in subjects' choices further we ran a majority rule vote experiment which provided complementary results to the observations in E&S and in B&O. First, since the payoff distributions achieved in cooperation games seldom feature perfect equality, as distribution A in B&O does, and since inequality aversion might have a greater influence on choices when it is the case<sup>55</sup>, we used a payoff matrix that allocates different amounts to the different player roles in all cases, similarly to E&S - in the two distributions the payoff to the different roles could be strictly ordered. Second, as in B&O and contrary to E&S we included a condition involving role certainty, where players knew which pair of alternative payoffs they were facing. In order to

<sup>55</sup> Güth et al. (2001), for instance, observed that unbalanced proposals in fixed-allocation ultimatum games tend to be chosen and accepted more often if equal distributions is not an available alternative.



collect subjects' general preferences between the two distributions, we used two conditions in which players did not know which roles they would be attributed and made one choice only to indicate their unique preferred distribution. We did not use the kind of strategy method applied in B&O, where subjects make one choice for each of the possible roles, since it does not address the same kind of preferences, nor the randomization procedure in S&O which similarly did not allow them to assess the general preferences of the subjects.

We ran three sessions involving 18 subjects each, none of them having previously participated in distribution game experiments. They were recruited through ORSEE from the subject pool of the LEES and each participated in one session only. In each condition, we formed different groups of three players, so that each subject would never be matched twice with the same other players. In each session all subjects played for three rounds, under the three different conditions. In order to insure independence between conditions we chose one of the rounds randomly to determine the payoffs for all subjects at the end of the experiment, we did not unveil the round outcome until the end of the last round and we varied the order in which the conditions were presented to subjects using all six possible combinations equally over the three sessions. Table 5.3 displays the payoff matrix that we used in all conditions. In each group, each subject had to voice their preferred distribution and the payoffs were then determined according to the distribution preferred by two or more subjects.

*Table 5.3: Game payoffs.*

Distribution	A	B
Player 1	18	33
Player 2	12	12
Player 3	10	5
Total	40	50

### 3.2. *Meeting John Rawls*

We thus used three different conditions where the information about the player's role varied. In addition to the certainty condition (Condition C for 'certainty'), players had also to choose between the same two distributions knowing that each role would be attributed to one player

of their group (Condition R for 'risk') and knowing only that themselves as well as the two other players could be in one of the three roles no matter the role attributed to the two other ones (Condition U for 'uncertainty'). This specificity of our design was implemented as an attempt to investigate the necessary conditions of procedural equity, a consideration which stems from the other motivation of our study, the much debated theory of justice proposed by John Rawls (1971). In this perspective, Conditions R and U can indeed be considered as two reduced implementations of the 'veil of ignorance', which is considered by Rawls as the necessary condition for the expression of the collective choice on the principles of justice themselves. In Rawls own words:

"The idea of the original position is to set up a fair procedure so that any principles agreed to will be just. The aim is to use the notion of pure procedural justice as a basis of theory. Somehow we must nullify the effects of specific contingencies which put men at odds and tempt them to exploit social and natural circumstances to their own advantage. Now in order to do this I assume that the parties are situated behind a veil of ignorance. They do not know how the various alternatives will affect their own particular case and they are obliged to evaluate principles solely on the basis of general considerations" (Rawls, 1971, p.136).

Though Rawls is setting out a general, fictitious procedure for society as a whole to choose its principles of justice, the 'original position' implies, among others, that none of the parties, the members of society, knows "his place in society, his class position or social status; nor does he know his fortune in the distribution of natural assets and abilities" (id., p.137). Within the framework of distribution games, the veil of ignorance can thus be translated into limited information conditions where subjects must choose only one of the alternative distributions without knowing what role they will be matched with. Given the broad scope of parties' ignorance of their fate in the original position as described by Rawls, the difference between conditions U and R can be seen as a small step from the most basic form of role uncertainty - where subjects know the group composition and thus the actual distribution of payoffs between the three members of the group - to a situation where subjects are also prevented from knowing the group composition - how many of each role will be represented.

### 3.3. *Maximin vs. 'efficiency'*

In the original position, behind the veil of ignorance, Rawls argues, it is rational for individuals to opt for the two principles of justice implied by his conception of 'justice as fairness' :

"First principle: each person is to have an equal right to the most extensive basic liberty compatible with a similar liberty for others.

Second principle: social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone's advantage, and (b) attached to positions and offices open to all." (ibid., p.60)

The second principle is later on given a more precise content, with the phrase 'everyone's advantage' specified as the 'difference principle' as opposed to the principle of efficiency, so that the new wording states that:

"Social and economic inequalities are to be arranged so that they are both (a) *to the greatest benefit of the least advantaged* and (b) attached to offices and positions open to all under conditions of fair equality of opportunity. " (ibid., p. 83, emphasis added)

Further in his essay Rawls introduces the maximin rule as a way to interpret the second principle. The principles themselves are of course presented as impartial views of the contractual structure of society, whereas the maximin rule is presented as a "heuristic device" describing an individual in a situation where they need to make a decision while "faced with several possible circumstances which he may or may not obtain" (ibid., p. 153). In the original position, Rawls argues, individuals would choose according to the maximin rule, for (i) "the situation is one in which a knowledge of likelihoods is impossible, or at best extremely insecure", which is required because the maximin rule "takes no account of the likelihoods of the possible circumstances" (ibid., p. 154), and (ii) the gains associated with the highest of the worst outcomes, which determines the solution, is high enough so that the person does not care about any higher gain given the possibly unacceptable outcomes otherwise faced.

The justification for the second principle, and for the choice of the maximin solution, as it appears, does not rely on equity or self-disinterested considerations. Among other elements

supporting the argument made by Rawls for the difference principle, the rationality of the parties when voicing their choice among alternative principles of justice in the original position is crucial. The motivations behind the choice are clearly based on self-interest, as applied when behind the veil of ignorance:

"The parties do not seek to confer benefits or to impose injuries on one another; they are not moved by affection or rancor. Nor do they try to gain relative to each other; they are not envious or vain. Put in terms of a game, we might say: *they strive for as high an absolute score as possible. They do not wish a high or a low score for their opponents, nor do they seek to maximize or minimize the difference between their successes and those of others.* The idea of a game does not really apply, since the parties are not concerned to win but to get as many points as possible judged by their own system of ends". (ibid., pp. 144-145, emphasis added)

This is a clear formulation of individualism as defined in the social value orientations literature<sup>56</sup>, by which agents are seeking to "maximize their own outcome without any concern for the partner's outcome" (Kollock, 1998). The maximin solution, in the rawlsian view, is thus not motivated by inequality aversion but by a form of what economic theory often denotes as 'selfishness'.

Rawls further underscores the higher burden put on agents by the utilitarian principle of efficiency, or Knight optimum, pointing that it "requires a greater identification with the interests of others than the two principles of justice" (ibid., p. 177). The argument follows:

"When the two principles are satisfied, [...] there is a sense defined by the difference principle in which everyone is benefited by social cooperation. [...] When the principle of utility is satisfied, however, there is no such assurance that everyone benefits, [...] thus the scheme will not be stable *unless those who must make sacrifices strongly identify with interests broader than their own*". Looking at the question from the standpoint of the original position, the parties recognize that *it would be highly unwise if not irrational to choose principles which may have consequences so extreme that they could not accept them in practice.* (ibid., pp. 177-178, emphasis added)

---

<sup>56</sup> See e.g. Kuhlman and Marshello (1975), McClintock and Liebrand (1988), Van Lange et al. (1992), and Kollock (1998) for a review.

As in Game II and III in B&O our distribution A is the maximin, the FFC and the ERC solution, whereas distribution B is the efficiency solution. Since the payoff for Player 2 is the same in both distributions, and since we kept the payoffs ordering between players, Player 1 is better off in terms of monetary gain in distribution B while Player 3 is better off in distribution A. This feature allows us to somehow refine the analysis of the weight the various trade-offs between the distributions that are diversely supported by the equity, efficiency and 'monetary selfish' solutions according to the player under consideration.

### 3.4. *Equity vs. 'selfishness'*

With equity aversion on the one side, as an other-regarding preference combination<sup>57</sup> of altruism (reducing favorable difference) and fairness (reducing unfavorable difference) operationalized by either of the definitions of ERC and FFC, selfishness is on the other side the usual label for the alternative motivation. Though it is widely used to define the solution which attributes the highest payoff to the concerned player it is not always clear what it is intended to capture on the motivational side, further than this support to the monetary payoff maximizing choice, when distributional concerns are involved. The outcome associated with the highest payoff for the player may for instance provide a narrower payoff difference for the other players, which would mean a non-selfish driver leading to a seemingly selfish outcome, and in any case players are more generally assumed not to be choosing from an impartial point of view but according to their very own preferences.

Table 5.4 summarizes the diverse outcomes as predicted by the behavioral drivers so far. Whereas the maximin, the ERC and the FFC predictions can be gathered in the case of this matrix<sup>58</sup>, maximin and efficiency solutions are by definition a group property while solutions based on self-favoring monetary concerns will be attributed to a different distribution according to the player concerned. We denote by 'pure selfish' the choice supported by purely individual motives, pointing to the distribution that provides the highest monetary payoff to the player. The 'competitive' solution, termed following the competitive orientation mentioned in Section 6.2 of Chapter 2, catches the highest total difference between the player's payoff and that of each of the other players.

<sup>57</sup> One could incidentally question to what extent these two components might constitute the two sides of the same behavioral driver. The difference is explicitly recognized in FCC, where a specific parameter is used according to the sign of the payoff difference, though the qualitative treatment is the same.

<sup>58</sup> It can indeed also be argued that the maximin solution is based on a form of inequality aversion principle, maybe more accurately described as an impartial preference for the least disadvantageous distribution from the point of view of the worst off among the players of the group.

*Table 5.4: Predicted solution, by player and overall.*

Solution	Maximin, ERC, FFC	Pure selfish	Efficiency	Competitive
Player 1	A	B	B	B
Player 2		-		A
Player 3		A		A
All	A	-	B	A

Though we intended to use alternative matrices where the 'equity' solutions would not all support distribution A, we chose to run first these three sessions using only one matrix in order to generate a first set of data where both efficiency and maximin were clearly marked and differentiated from other solutions for most players. This choice allowed us to set aside the potential problems of dependency arising from the order in which the different matrices would be proposed to the subjects, and to adjust future sessions according to the most salient features of the results from this first matrix.

**4. Results**

*4.1. Choosing a distribution: points victory for equity*

As shown in Table 5.5, the outcome of the majority rule was distribution A being chosen by 30 groups out of 54, or 56% of all cases. Both distributions were equally chosen under condition R whereas distribution A collected a one-group and a two-group majority in condition U and condition S respectively. It appears thus that neither of the two distributions is critically more attractive to players than the other, and that neither maximin and equity on the one side nor efficiency on the other side are strong enough drivers to cancel one another, even under uncertainty, especially when players are under diverging tensions from the point of view of the selfish and of the competitive choices, as it is the case in this matrix.

Table 5.5: Group choices, by condition.

Condition	S	R	U	Total	Total %
Distr. A	11	9	10	30	56%
Distr. B	7	9	8	24	44%
Total	18	18	18	54	100%

Table 5.6 displays the outcome of the individual votes, aggregated over the three sessions. The first indication is that distribution A gathered a slight majority of votes in condition S and R (54%), while the outcome was reversed in condition U. All three conditions taken together, the two distributions thus shared subjects' preferences almost evenly, showing that here again no one-sided theoretical prediction would prevail and fully explain the results.

Under condition S, among those who were concerned (Player 1 and Player 3), all subjects but one (97%) chose the 'pure selfish' distribution which allocated them the highest monetary payoff, and which in this matrix and for these players was also the 'competitive' solution. Among them, all subjects in the role of Player 1 chose thus the efficiency solution while those in the role of Player 3 overwhelmingly chose the maximin solution. This observation points to a seemingly predominant role of individual over group outcome considerations where the monetary payoffs to the player are at stake. It appears indeed that in this condition the group outcome was mostly determined by the choices of Player 2 subjects. Two thirds of the latter chose distribution A, which somehow questions the extent of the role of efficiency considerations and, though the null hypothesis of a random choice cannot be rejected (binomial test,  $p$ -value = 0.1189), supports the combined dominance of maximin and equity considerations when monetary payoffs are unchanged.

Concerning the effect of role uncertainty, which was introduced as an implementation of Rawls's veil of ignorance, individual choice counts exhibit no difference between condition S and condition R, with the two distributions sharing the votes in the same proportion in both conditions. It thus appears that role uncertainty *per se* is not enough either to deepen or to reverse the relative share of the two distributions and that the veil of ignorance may not be easily operationalized in experimental settings. Interestingly, the numbers were exactly reversed in condition U where subjects faced a higher uncertainty of relative payoffs within the group and where distribution B was chosen by a few more subjects. This also indicates

that procedural equity may play only a marginal role when stakes and procedures do not go beyond monetary payoffs.

*Table 5.6: Distribution of votes, by condition and broken down by player for condition S.*

Condition	S				R	U	Total	Total %
Player	1	2	3	Total	-	-	All	All
Distr. A	0	12	17	29	29	25	83	51%
Distr. B	18	6	1	25	25	29	79	49%
Total	18	18	18	54	54	54	162	100%

**4.2. The hidden effects of risk and uncertainty: intra-player analysis**

Though role uncertainty did not clearly seem to have much influence on players' choices at the aggregate level, looking at the three conditions separately, it is possible to investigate intra-player observations and see whether players consistently made the same choice across conditions so that differences at the aggregate level are only due to marginal switches, or if changing conditions triggered cross-switches by players having made opposite choices.

Table 5.7 displays for each of the three possible pairs of conditions the number of players who made the corresponding pair of choices. The second and fifth columns indicate consistent choices from one condition to the other while the third and fourth columns correspond to switches in opposite directions.

The figures tend to support a greater effect of role uncertainty on individual choices than the aggregated data would indicate. All player roles taken together, 37% of the subjects (20 out of 54) reversed their votes between condition S and condition R, 52% did so between condition S and condition U and 26% between conditions R and U. The first conclusion that can be drawn from these counts is thus that condition U actually triggered a more differentiated effect than condition R compared to the certainty condition. Moreover, the difference between the S/R and the S/U switches is mostly explained by Player 2 and Player 3 subjects switching from distribution A to distribution B, which triggered the reversal of votes in favor of distribution B observed in condition U, showing that equity and maximin considerations were less relevant in players' choices when the actual relative payoffs were unknown to players at the time of their choices. In that condition, subjects could indeed face



an egalitarian distribution yielding any of the three possible payoffs to the three group members.

*Table 5.7: Intra-player choices.*

S vs. R

Role in S	A-A	A-B	B-A	B-B	Total
Player 1	0	0	9	9	18
Player 2	9	3	0	6	18
Player 3	10	7	1	0	18
Total	19	10	10	15	54

S vs. U

Role in S	A-A	A-B	B-A	B-B	Total
Player 1	0	0	11	7	18
Player 2	6	6	0	6	18
Player 3	7	10	1	0	18
Total	13	16	12	13	54

R vs. U

Role in S	A-A	A-B	B-A	B-B	Total
Player 1	8	1	3	6	18
Player 2	6	3	0	9	18
Player 3	6	5	2	5	18
Total	20	9	5	20	54

## 5. Discussion

The above results strongly support that subjects' monetary payoffs play a first order role on their individual choices and further give some indications on outcome indecisiveness when these pure selfish motives are absent. However, some specificities of the design could be addressed.

First, the matrix used in these three sessions does not allow us to fully discriminate between alternative solutions for all player roles, so that our observations cannot all be directly connected to unique determinants. It could be relevant, for instance, to discriminate between the pure selfish and the competitive outcomes, the latter appearing to possibly play a non negligible role in players' choices though it has been somehow ignored previously, both in behavioral theories and experimental settings. Subjects had furthermore to choose between two distributions only, reducing thus their ability to fine tune their distribution choice and implying some framing effect. Falk et al. (2003) found for instance that unequal offers of (8,2) - allocating 8 to the proposer and 2 to the responder - were more often rejected if the alternative was (5,5) than if the alternative was (10,0). It would thus be worth using different payoff matrices, making perfect equality distribution available or varying the relative stakes to the players.

Second, our setting may be prone to the 'complicity effect' observed in Charness and Rabin (2002), Bolton and Zwick (1995), and first mentioned by Charness (2000) as the "responsibility-alleviation effect", a phenomenon which is described as occurring whenever "a shift of responsibility to an external authority dampens internal impulses towards honesty, loyalty, or generosity". In the same paper, Charness states that "in a gift-exchange experiment, we find that subjects respond with more generosity (higher effort) when wages are determined by a random process than when assigned by a third party, indicating that even a slight shift in perceived responsibility for the final payoffs can change behavior". In our experiment, since the chosen distribution is determined according to the combined votes of the three players in the group, each player may feel less compelled by equity or any other-regarding preferences. The choices of subjects in Player 1 role in condition S, for instance, may thus reflect only weakly their distributional concerns.

Third, though players were presented with the three conditions in all possible orders we could not completely rule out the existence of an ordering effect. Round by round breakdown of players' choices (see Table 5.8 below) indeed shows that all 6 subjects who were in Player 2 role when condition S was applied in the first round chose distribution A. Though we do not have enough observations to be able to reject the null hypothesis of a random distribution of player choices, a larger set of data would allow us to identify more decisively a possible preference for distribution B for players in that situation.

**Table 5.8:** *Distribution of votes in condition S broken down by round.*

Round	1				2				3				Total
	1	2	3	Total	1	2	3	Total	1	2	3	Total	
Distr. A	0	6	6	12	0	3	5	8	6	3	0	9	29
Distr. B	6	0	0	6	6	3	1	10	0	3	6	9	25
Total	6	6	6	18	6	6	6	18	6	6	6	18	54

Finally, only reduced versions of the veil of ignorance were operationalized through conditions R and U, for experimental settings can only involve choices between monetary distributions, not the determination of whole principles of justice to be applied to society and determine the broader life conditions of subjects. However, since Rawls explicitly suggested that the maximin rule was a good candidate to approximate the difference principle and even gave a detailed illustration of its application, our attempt to give an experimental account of its effectiveness is justified. In our two rawlsian conditions subjects actually knew probabilities and expected payoffs, though calculation was less obvious in condition U (and the result was not given) so that a more accurate test of the maximin rule as described by Rawls could be provided by a new condition where players would actually not be able to calculate expected payoffs, following his argument that "it must be that the situation is one in which a knowledge of likelihoods is impossible, or at best extremely insecure".

## 6. Conclusion

The experiment presented in this chapter had a twofold motivation, disentangling distributional concerns at the individual level on the one hand and testing alternative principles guiding collective choices on the other. The three experimental sessions run so far brought some clear cut results to the former as well as some preliminary indications to the latter.

Under certainty, all players chose the efficient distribution when it was also the one that allocated them the highest payoffs, regardless of equity concerns. Given that both distributions yielded some positive amounts to all players, however, it may be more precise to conclude that inequity aversion had no observable effect on the choices made by these players. It is indeed not to be excluded that distributional considerations may have been

taken into account but only to be dominated by the monetary gain for the player, since the relative share of the total outcome in efficient distribution between the highest and the lowest payoff was close to the average offered allocation observed in dictator game experiments. When players' payoffs were not affected by the chosen distribution, efficiency considerations alone were not enough to balance distributional concerns, though our sample was too small to achieve statistical significance on this point.

These observations however support that endogenous cooperation levels may be more related to the repetition of interactions between the same players than to altruism or equity considerations. This would also be more in line with the end effect observed in most public good games, which hardly fits inequity aversion models unless one assumes that equity concerns fade away for some reasons, but occurs precisely as the temporal horizon narrows down.

Contrastingly, role uncertainty leads to an almost even distribution of choices between the two alternatives. Again, our data provide support for a mitigated effect of inequity aversion, since the aggregated individual preferences were reversed when subjects also faced some uncertainty on the actual distributive features of the outcome - when an egalitarian distribution was a possible outcome - in condition U. However, in both conditions, risk aversion was not controlled for and may thus have also interfered with the other preferences.

Finally, at the collective level, preference heterogeneity entails that no decisive conclusion can be drawn from our settings on the choice between the maximin and the efficiency solution. This exploration will be taken further in forthcoming experimental sessions.



---

## GENERAL CONCLUSION

---

We will give here an overview of our main results, present a few illustrative empirical facts and suggest some possible future developments.

In line with our preliminary remarks on the diverging views on cooperation offered by the main approaches of collective production processes in economics, we concluded from the literature review presented in Chapter 1 that cooperation could be considered either as an emergent phenomenon or, more traditionally, as the outcome of constrained behavior. Broader behavioral theories emanating from neighboring fields allow us, however, to assume that none of the two approaches should be excluded and that efforts to integrate them into a common framework may well be rewarding on the long run. The broad range of experimental results reviewed in Chapter 2 seems to give some credit to the two alternative sets of behavioral assumptions. The heterogeneity of individual choices, which can be observed through the various dimensions of the experimental settings, clearly points to the necessity of upholding the various behavioral drivers conjectured by the competing theories.

Instancing the limitations of the third assumption of contract theories (we refer here and below to the five assumptions detailed in Chapter 1, Section 8), group outcome and distributional concerns appear to be major determinants of the choices made by some players while others seem to choose according to their own monetary payoffs exclusively. Results from public good game experiments, including our own results presented in Chapter 3, further show that some players do not rule out the possibility of positive contributions from others, even when they themselves would chose not to contribute, which furthermore supports an extension of the fifth assumption. The gift-exchange game experiment in Chapter 4 provides additional evidence on the effect of repeated interactions between the same players. We observed that both wage and effort levels were higher in that case than in a stranger setting, and the highest in trust-based contracts, a result than cannot be explained if all agents were to choose myopically as stated in the fourth assumption. However, as the results of the distribution game experiment reported in Chapter 5 epitomize, agents are far from losing sight of their monetary interest, at least within the restricted area of laboratory experiments.

Taken together, our findings suggest a twofold conception of individual choices, cooperation and collective production processes as well, involving both short-term, monetary oriented, constrained contribution choices and long-term, achievement oriented, trust-based participation choices. Depending on the concerned agent, on the implications of the choice at stake and on the broader features of the situation, one or the other of these two components may dominate.

Though we deliberately stayed within the boundaries of the theoretical and experimental realms, numerous empirical illustrations of the emergence of cooperation can be found. We would briefly mention here the three following cases: producer cooperatives, cognitive communities and open-source software developers. Producer cooperatives illustrate the achievement of the cooperative outcome in a way that contract theories would typically rule out. Since capital ownership and employment are closely, if not exclusively, connected, the owners-employees face a real case public-good game. Studies in the field<sup>59</sup> show that this organizational form not only subsists as an alternative contractual arrangement, but also that their specific organizational goals increase their responsiveness to environmental changes. As a rough indication of their present extension, cooperative firms hire respectively 5% and 3.5% of the workforce<sup>60</sup> in Italy and in France. Incidentally, these relatively low numbers might call for closer inquiry.

The various cognitive communities mentioned in Chapter I, Section 5, play a crucial role in knowledge creation processes and provide another clear example of spontaneous contributions from agents engaged in productive activities, in a situation where incentive schemes would be hardly effective. Similarly, open-source software is mainly produced by large communities of developers, often strangers to each others, outside any other contractual arrangement than the licenses that grant open access to the source code. The extremely wide range of applications available under public license terms speaks for the success of these cooperation structures. These two last examples also illustrate the complementarity between the incentive-based and the trust-based components of collective production processes.

---

<sup>59</sup> See e.g. Bonin et al. (1993) for a synthesis of both theoretical and empirical studies, and yearly updates available on the website of the International Co-operative Alliance, [www.ica.coop](http://www.ica.coop).

<sup>60</sup> Source: Coop FR (The representative organization of the French co-operative movement, [www.entreprises.coop](http://www.entreprises.coop)).

From the present standpoint several research avenues can be considered.

Empirical studies could provide material for a better understanding of the general and specific factors underlying the determination of organizational forms between capitalist and cooperative firms. Since no decisive argument arise from either theoretical or experimental finding, the question remains open to empirical examination which would help uncovering factors that may have been ignored so far.

Throughout our experimental work some promising designs have been let aside which may further enlighten the individual cooperative choices, as well as the relative performance of alternative interaction structures. For instance, it has emerged from various discussions of the public-good experiment presented in Chapter 3 that voluntary participation could be implemented in such a way that both group size and multiplier remain constant. The second experiment included in this thesis (Chapter 4) could also be extended to involve more than two players and to let players choose the type of contractual arrangement itself, between a principal-agent and a public-good-like structure.

Finally, further work is necessary to advance the rough dichotomy observed in experimental results and identified in theoretical approaches as well, in order to refine the analysis of the interplay between these two components of cooperative behavior. The subtle dynamics of the participation and contribution of the members of the organization may well be accurately grasped only through the successful integration of the two approaches.





## APPENDICES

### Appendix A: Instructions to subjects - Chapter 3

---

#### *Bonjour et bienvenue*

Vous allez participer à une expérience de sciences sociales. Si vous suivez scrupuleusement ces instructions, les choix que vous ferez vous permettront de gagner une certaine somme d'argent. Toutes vos réponses seront recueillies à travers un réseau informatique et seront traitées de façon anonyme.

Il vous est instamment demandé de **ne pas communiquer avec les autres participants**. Si vous avez des questions concernant ces intructions, veuillez lever la main lorsque vous y serez invité(e). Vous indiquerez directement vos choix à l'ordinateur devant lequel vous serez assis(e) et celui-ci vous informera de vos gains réalisés au fur et à mesure du déroulement de l'expérience.

Les gains que vous réaliserez dépendent à la fois des décisions que vous prendrez et des décisions prises par les quatre autres membres de votre groupe. Ces gains seront comptabilisés en points. Votre total de points gagnés pendant l'expérience sera converti, à la fin de l'expérience, en euros, et cette somme vous sera versée. La procédure de conversion des points en euros est détaillée à la fin des instructions.

#### *Cadre général de l'expérience*

15 personnes participent à cette expérience. Les 15 personnes sont réparties en trois groupes de 5 personnes. Vous êtes donc membre d'un de ces groupes de **5 personnes**. Vous ne connaîtrez pas l'identité des autres personnes de votre groupe, ni pendant l'expérience, ni après. Au cours de l'expérience, vous n'interagirez qu'avec les quatre autres personnes de votre groupe et jamais avec les 10 autres personnes participant à l'expérience.

L'expérience comporte **20 périodes**.

### ***Déroulement d'une période.***

Au début de chaque période, chaque membre de votre groupe, y compris vous-même, dispose d'une dotation de **10 jetons**. Cette dotation est la même quels que soient les choix faits lors des périodes précédentes.

#### **Chaque période comporte deux étapes:**

##### **Première étape :**

Durant cette étape, vous avez la possibilité de participer à un projet commun.

- Si vous choisissez de ne pas participer à ce projet, la période sera terminée pour vous. Votre gain pour la période sera alors de 16 points (les 10 points provenant de vos jetons plus 6 points de revenu).
- Si vous avez choisi de participer, le nombre de participants au projet vous sera indiqué dès que tous les membres de votre groupe auront fait leur choix. Vous passerez alors à la deuxième étape.

##### **Deuxième étape :**

Durant cette étape, vous devez choisir votre contribution au projet. Vous devez répartir votre dotation entre les jetons que vous conservez, et ceux que vous attribuez au projet.

Le gain total (nombre de points) que vous réalisez est composé de deux parties:

- un gain qui provient des jetons conservés: chaque jeton conservé vous rapporte un point.
- un gain qui provient des revenus du projet, ce revenu est égal à 0,75 fois la somme des contributions des membres de votre groupe participant au projet, sauf si vous êtes le seul participant. Il est donc égal à:

$$0,75 \times \text{Somme des contributions de tous les participants.}$$

Un tableau vous a été remis qui vous indique quel sera votre **gain total** (revenu des jetons conservés + revenu du projet) en fonction du nombre de participants, de votre contribution (nombre de jetons que vous avez choisi d'attribuer au projet) et de la somme des contributions des autres participants (nombre total de jetons que les autres participants ont attribué au projet):

- Sélectionnez dans la première colonne du tableau le nombre de participants au projet (y compris vous-même) qui vous a été indiqué à la fin de la première étape.
- Sélectionnez ensuite la colonne correspondant à votre contribution au projet et la ligne correspondant à la contribution totale des autres participants. L'intersection de cette colonne et de cette ligne vous indique le gain total (nombre de points obtenus) que vous réaliserez au cours d'une période si ces contributions sont réellement choisies.

Cas particulier: Si vous êtes le seul participant, vous conservez simplement votre dotation initiale. Votre gain pour la période est donc de 10 points.

### **Exemples indicatifs:**

*Exemple 1: Supposez que lors de la première étape, vous choisissiez d'accepter la proposition de participer au projet et que votre écran vous indique alors que 4 membres de votre groupe ont choisi de participer (vous et trois autres membres de votre groupe).*

*La deuxième étape commence alors, et il vous est demandé de choisir combien de jetons vous allouez au projet commun. Supposez que vous choisissiez de contribuer à hauteur de 6 jetons. Supposez que dans le même temps, les autres participants aient choisi d'investir 18 jetons au total (vous ne connaîtrez ce montant qu'à la fin de la période). Votre gain total serait alors de 22 jetons (4 points de revenu des jetons que vous avez conservés plus  $0,75 \times (18 + 6) = 18$  points correspondant à votre part du projet commun).*

*Exemple 2: Supposez que lors de la première étape, vous choisissiez de refuser la proposition de participer au projet. Vous n'avez alors plus de décision à prendre pour cette période. Lorsque tous les autres membres de votre groupe ont fait leur(s) choix, un écran s'affiche. Cet écran vous indique le nombre de participants au projet, la contribution totale au projet, ainsi que votre gain de 16 points (les 10 points de votre dotation plus 6 points de revenu).*

Vous ne connaîtrez les décisions des autres participants que lorsque la période sera terminée. Lorsque tous les membres de votre groupe ont fait leur(s) choix, le revenu du projet est calculé ainsi que le gain de chaque participant. Vous êtes alors informé(e) de votre gain pour la période et de la contribution totale de votre groupe au projet. La période suivante commence dès que tous les membres de votre groupe sont prêts.

### ***L'historique.***

A tout moment, vous pouvez afficher l'historique du jeu de votre groupe. Celui-ci vous rappelle, pour chaque période passée: votre décision de participer ou non au projet commun, le nombre de participants au projet commun (de 1 à 5 sur les 5 membres de votre groupe, y compris vous-même si vous avez choisi de participer), votre contribution au projet si vous avez choisi de participer, la contribution totale de votre groupe au projet, votre gain total pour la période et la somme de vos gains depuis la première période.

### ***Les gains.***

Lorsque la 20<sup>e</sup> période est achevée, l'ordinateur affiche le montant total de vos gains en points cumulés au cours des 20 périodes. Ce montant total est la somme des points gagnés à chacune des 20 périodes. Il est ensuite converti en euros. Le facteur de conversion est de 1 euro pour 20 points ; autrement dit, un point vaut 0,05 euros. *Par exemple, si vos gains pour l'ensemble de l'expérience s'élèvent à 300 points, vous percevrez un montant de 15 euros en liquide.*

Avant de démarrer l'expérience, vous devrez répondre à un questionnaire posé par l'ordinateur afin de vérifier votre bonne compréhension des instructions

A la fin de l'expérience, un expérimentateur vous appellera individuellement pour vous permettre de percevoir votre gain en Euros.

**Merci pour votre participation**

---

**Appendix B:** Instructions to subjects - Chapter 4

---

***Bonjour, et bienvenue dans notre laboratoire.***

Vous allez participer à une expérience de sciences sociales. Si vous suivez scrupuleusement ces instructions, vos propres choix vous permettront de gagner une certaine somme d'argent. Toutes vos réponses seront recueillies à travers un réseau informatique et seront traitées de façon anonyme.

Il vous est instamment demandé de **ne pas communiquer avec les autres participants**. Si vous avez des questions concernant ces instructions, veuillez lever la main lorsque vous y serez invité(e). Vous indiquerez directement vos choix à l'ordinateur devant lequel vous serez assis(e) et celui-ci vous informera de vos gains réalisés au fur et à mesure du déroulement de l'expérience.

Les gains que vous réaliserez dépendent à la fois des décisions que vous prendrez et des décisions prises par votre partenaire. Ces gains seront comptabilisés en points. Votre total de points gagnés pendant l'expérience sera converti, à la fin de l'expérience, en euros, et cette somme vous sera versée. La procédure de conversion des points en euros est détaillée à la fin des instructions.

***Cadre général de l'expérience***

20 personnes participent à cette expérience. Vous serez répartis en 10 binômes. Chaque binôme est composé d'un joueur A et d'un joueur B. Vous serez donc soit le joueur A, soit le joueur B de l'un de ces dix binômes, selon ce que l'ordinateur devant lequel vous serez assis(e) vous indiquera. Vous ne connaîtrez pas l'identité de votre partenaire, ni pendant l'expérience, ni après. Au cours de l'expérience, vous n'interagirez qu'avec votre partenaire et jamais avec les 18 autres personnes participant à l'expérience.

L'expérience comporte **10 périodes**.

***Déroulement d'une période.***

- Au cours de chaque période, le joueur A puis le joueur B doivent prendre un certain nombre de **décisions**.

Dans un premier temps le joueur A choisit un des deux contrats proposés.

S'il choisit le contrat **1**, il choisit ensuite:

- le niveau d'effort  $e^*$  qu'il souhaite que le joueur B fournisse, compris entre 1 et 10,
- un montant  $w$  à verser au joueur B

S'il choisit le contrat **2**, il choisit ensuite:

- le niveau d'effort  $e^*$  qu'il souhaite que le joueur B fournisse, compris entre 1 et 10,
- un montant  $w$  à verser au joueur B,
- le montant de la pénalité  $f$  que le joueur B devra lui verser avec une probabilité de 1/3 au cas où il aurait choisi un niveau d'effort inférieur au niveau demandé par le joueur A.

Dans un deuxième temps, les choix du joueur A sont transmis au joueur B et celui-ci peut alors, soit refuser l'offre du joueur A, soit l'accepter:

- si le joueur B refuse l'offre du joueur A, aucun des deux joueurs ne reçoit de point pour la période en cours et la période suivante commence,
- si le joueur B accepte l'offre du joueur A il doit choisir un niveau d'effort  $e$ .

- Si le joueur A a choisi le contrat **1** et que le joueur B a accepté son offre, les **gains** réalisés par chacun des joueurs sont calculés de la manière suivante:

- le joueur A reçoit dix fois le niveau d'effort  $e$  choisi par le joueur B, moins le montant  $w$  versé au joueur B,
- le joueur B reçoit le montant  $w$  versé par le joueur A, moins le coût correspondant au niveau d'effort choisi  $c(e)$ .

Si le joueur A a choisi le contrat **2** et que le joueur B a accepté son offre et a choisi un niveau d'effort  $e$  supérieur ou égal au niveau d'effort demandé  $e^*$ , les **gains** réalisés par chacun des joueurs sont calculés ainsi:

- le joueur A reçoit dix fois le niveau d'effort  $e$  choisi par le joueur B, moins le montant  $w$  versé au joueur B, moins un coût de contrôle  $k$  de 10,
- le joueur B reçoit le montant  $w$  versé par le joueur A, moins le coût correspondant au niveau d'effort choisi  $c(e)$ .

Si le joueur A a choisi le contrat **2** et que le joueur B a accepté son offre et a choisi un niveau d'effort  $e$  inférieur au niveau d'effort demandé  $e^*$ , un tirage au sort détermine si le joueur B doit verser au joueur A la pénalité  $f$ . La probabilité pour que la pénalité soit versée est de  $1/3$ . Les **gains** réalisés par chacun des joueurs sont alors calculés ainsi:

- Le joueur A reçoit dix fois le niveau d'effort  $e$  choisi par le joueur B, moins le montant  $w$  versé au joueur B, moins le coût de contrôle  $k$  de 10, plus la pénalité  $f$  si le tirage au sort l'a déterminé.
  - Le joueur B reçoit le montant  $w$  versé par le joueur A, moins le coût correspondant au niveau d'effort choisi  $c(e)$ , moins la pénalité  $f$  si le tirage au sort l'a déterminé.
- Le montant  $w$  que le joueur A propose de verser au joueur B doit être supérieur ou égal au coût  $c(e^*)$  correspondant au niveau d'effort demandé  $e^*$ .

Le tableau suivant vous indique le coût associé à chacun des niveaux d'effort, quel que soit le contrat choisi:

<b>e</b>	1	2	3	4	5	6	7	8	9	10
<b>c(e)</b>	0	1	2	4	6	8	10	13	16	20

Si le joueur A choisit le contrat 2, il doit fixer un montant inférieur ou égal à 13 pour la pénalité  $f$ .

▪ **Exemples indicatifs:**

*Exemple 1:* Supposez que le joueur A choisisse le contrat 1, propose de verser un montant  $w$  de 20 au joueur B et lui demande de fournir un effort  $e^*$  de 5. Supposez que le joueur B accepte cette offre et choisisse un niveau d'effort  $e$  de 5.

Le gain du joueur A serait alors de:  $10 \times e - w = 10 \times 5 - 20 = 30$ .

Le gain du joueur B serait de:  $w - c(e) = 20 - c(5) = 20 - 6 = 14$ .



*Exemple 2:* Supposez que le joueur A choisisse le contrat 2, propose de verser un montant  $w$  de 20 au joueur B, lui demande de fournir un effort  $e^*$  de 10 et choisisse un montant de 10 pour la pénalité  $f$ . Supposez que le joueur B accepte cette offre et choisisse un niveau d'effort  $e$  de 5, inférieur au niveau d'effort demandé  $e^*$ .

Si le tirage au sort détermine que la pénalité doit être payée par le joueur B le gain du joueur A serait alors de  $10 \times e - w - k + f = 10 \times 5 - 20 - 10 + 10 = 30$ .

Le gain du joueur B serait alors de  $w - c(e) - f = 20 - c(5) - 10 = 20 - 6 - 10 = 4$ .

Si le tirage au sort détermine que la pénalité ne doit **pas** être payée le gain du joueur A serait alors de  $10 \times e - w - k = 10 \times 5 - 20 - 10 = 20$ .

Le gain du joueur B serait alors de  $w - c(e) = 20 - 6 = 14$ .

### ***L'historique.***

A tout moment, vous pouvez afficher l'historique du jeu de votre binôme. Celui-ci vous rappelle, pour chaque période **passée**:

- le contrat choisi,
- le montant  $w$  proposé par le joueur A,
- le niveau d'effort  $e^*$  demandé par le joueur A,
- le montant de la pénalité  $f$  choisi par le joueur A s'il a choisi le contrat 2,
- le coût de contrôle  $k$ , toujours égal à 10 (coût de contrôle déduit des gains du joueur A s'il a choisi le contrat 2),
- le choix du joueur B d'accepter ou non le contrat,
- le niveau d'effort  $e$  choisi par le joueur B,
- le coût  $c(e)$  correspondant au niveau d'effort  $e$  choisi par le joueur B,
- le montant de la pénalité  $f$  versée par le joueur B au joueur A le cas échéant,
- votre gain pour la période, en nombre de points,
- vos gains cumulés depuis la première période, en nombre de points.

### ***Les gains cumulés.***

Lorsque la 10<sup>e</sup> période est achevée, l'ordinateur affiche le montant total de vos gains en points cumulés au cours des 10 périodes. Ce montant total est la somme des points gagnés à chacune des 10 périodes. Il est ensuite converti en euros. Le facteur de conversion est de 1

euro pour 10 points: un point vaut 0,1 euro. *Par exemple, si vos gains pour l'ensemble de l'expérience s'élèvent à 150 points, vous percevrez un montant de 15 euros en liquide.*

Avant de démarrer l'expérience, vous devrez répondre à un questionnaire posé par l'ordinateur afin de vérifier votre bonne compréhension de ces instructions.

A la fin de l'expérience, un expérimentateur vous appellera individuellement pour vous remettre votre gain en euros.

### **Merci de votre participation**

### ***COMMENTAIRES***

Merci d'utiliser cette page pour nous faire part de vos commentaires sur le déroulement de l'expérience, sur la façon dont vous avez fait vos choix et d'une façon générale sur ce que vous avez pensé de l'expérience.

---

**Appendix C: Instructions to subjects - Chapter 5****Bonjour, et bienvenue dans notre laboratoire.**

*Vous allez participer à une expérience de sciences sociales. Si vous suivez scrupuleusement ces instructions, vos propres choix vous permettront de gagner une certaine somme d'argent. Toutes vos réponses seront recueillies à travers un réseau informatique et seront traitées de façon anonyme.*

*Il vous est instamment demandé de ne pas communiquer avec les autres participants. Si vous avez des questions concernant ces instructions, veuillez lever la main lorsque vous y serez invité(e). Vous indiquerez directement vos choix à l'ordinateur devant lequel vous serez assis(e) et celui-ci vous informera de vos gains.*

*Les gains que vous réaliserez dépendent à la fois des décisions que vous prendrez et des décisions prises par vos partenaires. Ces gains seront comptabilisés en points. Les points gagnés pendant l'expérience seront convertis, à la fin de l'expérience, en euros, et cette somme vous sera versée.*

***Cadre général de l'expérience***

18 personnes participent à cette expérience. Vous ne connaîtrez pas l'identité de vos partenaires, ni pendant l'expérience, ni après. Au cours de cette expérience, vous ne jouerez jamais plus d'une fois avec le même joueur. L'expérience comporte trois périodes.

***Déroulement d'une période***

Au début de chaque période, des **groupes de 3 joueurs** sont formés de sorte que chaque joueur ne joue jamais plus d'une fois avec les mêmes joueurs. Au cours d'une période, chaque joueur interagit uniquement avec les autres joueurs de son groupe.

Chaque groupe est composé d'un **joueur 1**, d'un **joueur 2** et d'un **joueur 3**. Les rôles 1, 2 et 3 sont attribués aléatoirement. Chaque groupe doit choisir entre l'**option 1** et l'**option 2**, chacune correspondant à des gains différents pour certains joueurs du groupe. Le nombre de points attribués par les options 1 et 2 aux trois joueurs sont indiqués dans le tableau suivant:

	<b>Option 1</b>	<b>Option 2</b>
<b>Joueur 1</b>	18	33
<b>Joueur 2</b>	12	12
<b>Joueur 3</b>	10	5
<b>Total</b>	<b>40</b>	<b>50</b>

Vous devrez indiquer votre **préférence** pour l'une ou l'autre des options en cliquant dans la case correspondante.

Au moment d'indiquer votre préférence, vous serez dans l'une des trois situations suivantes :

1. **Situation A** : vous **connaissez** votre rôle. Vous êtes soit le joueur 1, soit le joueur 2, soit le joueur 3.
2. **Situation B** : vous **ne connaissez pas** votre rôle. Celui-ci vous sera attribué de manière aléatoire par l'ordinateur. En revanche, les **chances que vous avez** d'être le joueur 1, le joueur 2 ou le joueur 3 sont **connues** : vous **savez** que vous avez **une chance sur trois** d'être le joueur 1, le joueur 2 ou le joueur 3.
3. **Situation C** : vous **ne connaissez pas** votre rôle. Celui-ci vous sera attribué de manière aléatoire par l'ordinateur. De plus, les **chances que vous avez** d'être le joueur 1, le joueur 2 ou le joueur 3 sont **inconnues** : vous **ne savez pas** selon quelle probabilité vous serez le joueur 1, le joueur 2 ou le joueur 3.

La situation dans laquelle votre groupe se trouve vous sera indiquée au début de la période.

Lorsque tous les joueurs du groupe ont indiqué leurs préférences, **l'option choisie est déterminée selon la règle majoritaire**. L'option choisie est donc l'option préférée par au moins deux joueurs du groupe.

### *Les gains*

Lorsque la dernière période est terminée les résultats des trois périodes s'affichent sur votre écran.

Les gains sont déterminés de la façon suivante: l'une des trois périodes est tirée au sort et le gain que vous avez réalisé lors de cette période est votre gain pour l'expérience.

Les gains sont alors convertis en euros à parité égale (le facteur de conversion est de un euro pour un point).

*Avant de démarrer l'expérience, vous devrez répondre à un questionnaire posé par l'ordinateur afin de vérifier votre bonne compréhension de ces instructions.*

*A la fin de l'expérience, un expérimentateur vous appellera individuellement pour vous remettre votre gain en euros.*

**Merci pour votre participation.**

### **COMMENTAIRES**

Merci d'utiliser cette page pour nous faire part de vos commentaires sur le déroulement de l'expérience, sur les raisons de vos choix et d'une façon générale sur ce que vous avez pensé de l'expérience.

Nous allons vous appeler individuellement par numéro de poste pour vous remettre la somme que vous avez gagnée.

**Merci et à bientôt!**

—

## REFERENCES

- Ahn, T.H., R.M. Isaac and T.C. Salmon (2008)**, "Endogenous group formation", *Journal of Public Economic Theory*, 10(2), pp. 171-194.
- Alchian and Demsetz (1972)**, "Production, Information Costs and Economic Organization", *American Economic Review*, 62(5), pp. 777-795.
- Andreoni, J. (1988)**, "Why free ride? Strategies and learning in public goods experiments", *Journal of Public Economics*, 37(3), pp. 291-304.
- Argyris, C. and D. Schön (1978)**, *Organizational Learning: A Theory of Action Perspective*, Addison Wesley, Reading.
- Aristote (1953)**, "De la Mémoire et de la Réminiscence", in Aristote, *Petits Traités d'Histoire Naturelle*, Belles Lettres, Paris, [around 330 B.C. ?].
- Aristote (1992)**, *Ethique de Nicomaque*, Flammarion, Paris [lecture notes: 325-322 B.C. ?].
- Ashley, R., S. Ball and C. Eckel (2002)**, "Analysis of experimental public good data using dynamic panel regression models", Working Paper, Dept. of Economics, Virginia Tech.
- Axelrod, R. And W.D. Hamilton (1981)**, "The Evolution of Cooperation", *Science*, 211 (4489), pp. 1390-1396.
- Bardsley, N. and R. Sausgruber (2005)**, "Conformity and reciprocity in public good provision", *Journal of Economic Psychology*, 26(5), pp. 664-681.
- Barnard, C.I. (1938)**, *The Functions of the Executive*, Harvard University Press, Cambridge.
- Barnard, C.I. (1948)**, *Organization and Management: Selected Papers*, Harvard University Press, Cambridge.
- Barnard, C.I. (1958)**, "Elementary conditions of business morals", *California Management Review*, 1(1), pp. 1-13.
- Beckenkamp, M., H. Hennig-Schmidt and F.P. Maier-Rigaud (2007)**, "Cooperation in symmetric and asymmetric prisoner's dilemma games", Working Paper Series of the Max Planck Institute for Research on Collective Goods, Bonn, 2006/25.
- Bochet, O., T. Page and L. Putterman (2006)**, "Communication and punishment in voluntary contribution experiments", *Journal of Economic Behavior and Organization*, 60(1), pp. 11-26.

- Bolton, G. E. and A. Ockenfels (2000)**, "ERC: a theory of equity, reciprocity and competition", *American Economic Review*, 90(1), pp. 166-193.
- Bolton, G. E. and A. Ockenfels (2006)**, "Inequality aversion, efficiency and maximin preferences in simple distribution experiments: comment", *American Economic Review*, 96(5), pp. 1906-1991.
- Bolton, G. E. and R. Zwick (1995)**, "Anonymity versus punishment in ultimatum bargaining", *Games and Economic Behavior*, 10(1), pp. 95-121.
- Bonin, J.P., D.C. Jones and L. Putterman (1993)**, "Theoretical and Empirical Studies of Producer Cooperatives: Will Ever the Twain Meet?", *Journal of Economic Literature*, 31(3), pp. 1290-1320.
- Bonacich, P., G. Shure, J. Kahan and R. Meeker (1976)**, "Cooperation and group size in the n-person prisoners' dilemma", *Journal of Conflict Resolution*, 20(4), pp. 687-706.
- Bowles, S. and H. Gintis (2004)**, "The evolution of strong reciprocity: cooperation in heterogeneous populations", *Theoretical Population Biology*, 65(1), pp. 17-28.
- Brown, J.S. and P. Duguid (1991)**, "Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation", *Organization Science*, 2(1), pp. 40-57.
- Buchanan, J.M. (1965)**, "An Economic Theory of Clubs", *Economica*, 32(1), pp. 1-14.
- Buckley, E. and R. Croson (2006)**, "Income and wealth heterogeneity in the voluntary provision of linear public goods", *Journal of Public Economics*, 90(4-5), pp. 935-955.
- Burlando, R.M. and F. Guala (2005)**, "Heterogeneous agents in public good experiments", *Experimental Economics*, 8(1), pp. 53-54.
- Charness, G. (2000)**, "Responsibility and effort in an experimental labor market", *Journal of Economic Behavior and Organization*, 42(3), pp. 375-384.
- Charness, G. and M. Rabin (2002)**, "Understanding Social Preferences with Simple Tests", *Quarterly Journal of Economics*, 117(3), pp. 817-869.
- Cherry, T.L., S. Kroll and J.F. Shogren (2005)**, "The impact of endowment heterogeneity and origin on public good contributions: evidence from the lab", *Journal of Economic Behavior and Organization*, 57(3), pp. 357-365.

- Chiappori, P. A. and B. Salanie (2003)**, "Testing contract theory: a survey of some recent work", in M. Dewatripont, L. Hansen, and P. Turnovsky (eds.), *Advances in Economics and Econometrics: Theory and Application, Eighth World Congress*, Cambridge University Press, Cambridge, pp. 115-149.
- Cialdini, R. (1993)**, *Influence: Science and Practice*, Harpers Collins, New York.
- Cinyabuguma, M., T. Page and L. Putterman (2005)**, "Cooperation under the threat of expulsion in a public goods experiment", *Journal of Public Economics*, 89(8), pp. 1421-1435.
- Clark, K. and M. Sefton (2001)**, "The sequential prisoner's dilemma: evidence on reciprocation", *The Economic Journal*, 111(1), pp.51-68.
- Cohendet, P. and P. Llerena (2003)**, "Routines and incentives: the role of communities in the firm", *Industrial and Corporate Change*, 12(2), pp. 271-297.
- Cook, K.S. And R. Hardin (2001)**, Norms of Cooperativeness and Networks of Trust, in M. Hechter and K.D. Opp (eds.), *Social Norms*, Sage, New York, pp. 327-347.
- Coricelli, G., D. Fehr and G. Fellner (2004)**, "Partner selection in public goods experiments", *Journal of Conflict Resolution*, 48(3), pp. 356-378.
- Croson, R. (1996)**, "Partners and strangers revisited", *Economics Letters*, 53(1), pp. 25-32.
- Croson, R. and M.B. Marks (2000)**, "Step returns in threshold public goods: a meta- and experimental analysis", *Experimental Economics*, 2(3), pp. 239-259.
- Cyert, R.M. And J.G March (1963)**, *A Behavioral Theory of the Firm*, Prentice-Hall, Engelwood Cliffs.
- Dawes, R.M., J. McTavish and H. Shaklee (1977)**, "Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation", *Journal of Personality and Social Psychology*, 35(1), pp. 1-11.
- Dawes, R.M., J.M. Orbell, R. Simmons and A.J.C. Van de Kragt (1986)**, "Organizing groups for collective action", *American Political Science Review*, 80(4), pp. 1171-1185.
- Deci, E.L. (1971)**, "Effects of externally mediated rewards on intrinsic motivation", *Journal of Personality and Social Psychology*, 18(1), pp. 105-115.
- Deci, E.L. and R.M. Ryan (1980)**, "The empirical exploration of intrinsic motivational processes", in L. Berkowitz (ed.), *Advances in Experimental Social Psychology 13*, Academic Press, NewYork, pp. 39-80.



- Deci, E.L. and R.M. Ryan (1985)**, *Intrinsic Motivation and Self-Determination in Human Behavior*, Plenum Press, New York.
- Deci, E.L. and R.M. Ryan (1999)**, "The support of autonomy and the control of behavior", *Journal of Personality and Social Psychology*, 53(6), pp. 1024-1037.
- Deci, E.L. and R.M. Ryan (2000)**, "The "what" and "why" of goal pursuits: human needs and the self-determination of behavior", *Psychological Inquiry*, 11(4), pp. 227-268.
- Dickinson, D. and M.-C. Villeval (2008)**, "Does monitoring decrease work effort? The complementarity between agency and crowding-out theories ", *Games and Economic behavior*, 63(1), pp. 56-76.
- Dosi, G., D. Levinthal and L. Marengo (2003)**, "Bridging contested terrain: linking incentive-based and learning perspectives on organizational evolution", *Industrial and Corporate Change*, 12(2), pp. 413-436.
- Dosi, G. and L. Marengo (1994)**, "Toward a theory of organizational competencies", in R.W. England (ed.), *Evolutionary Concepts in Contemporary Economics*, Michigan University Press, Ann Arbor, pp. 157-178.
- Durckheim, E. (1973)**, *De la Division du Travail Social*, PUF, Paris [1893].
- Ehrhart, K. and C. Keser (1999)**, "Mobility and cooperation: On the run", Working Paper, Cyrano Scientific Series, 99s-24.
- Engel, C. (2010)**, "Dictator game: a meta study", Working Paper Series of the Max Planck Institute for Research on Collective Goods, Bonn, 2010/07.
- Engelmann, D. and M. Strobel (2004)**, "Inequality aversion, efficiency and maximin preferences in simple distribution experiments", *American Economic Review*, 94(4), pp. 857-869.
- Erev, I. and A. Rapoport (1990)**, "Provision of step-level public goods: the sequential contribution mechanism", *Journal of Conflict Resolution*, 34(3), pp. 401-425.
- Ertan, A., T. Page and L. Putterman (2005)**, "Can endogenously chosen institutions mitigate the free-rider problem and reduce perverse punishment?", Working Paper, Brown University.
- Falk, A., E. Fehr and U. Fischbacher (2003)**, "On the nature of fair behavior", *Economic Inquiry*, 41(1), pp. 20-26.

- Falk, A. and U. Fischbacher (2006)**, "A theory of reciprocity", *Games and Economic Behavior*, 54(2), pp. 293-315.
- Falk, A., S. Gächter and J. Kovács (1999)**, "Intrinsic motivations and extrinsic incentives in a repeated game with incomplete contracts", *Journal of Economic Psychology*, 20(3), pp. 251-284.
- Farrell, J. and M. Rabin (1996)**, "Cheap talk", *Journal of Economic Perspectives* 10(3), pp. 103-118.
- Fehr, E., U. Fischbacher and S. Gächter (2002)**, "Strong reciprocity, human cooperation and the enforcement of social norms", *Human Nature*, 23(1), pp. 1-25.
- Fehr, E. and S. Gächter (1998)**, "How effective are trust- and reciprocity-based incentives?", in A. Ben-Ner and L. Putterman (eds.), *Economics, Values and Organizations*, Cambridge University Press, Cambridge, pp. 337-363.
- Fehr, E. and S. Gächter (2000)**, "Cooperation and punishment in public good experiments", *American Economic Review*, 90(4), pp. 980-994.
- Fehr, E. and S. Gächter (2002)**, "Do incentive contracts undermine voluntary cooperation?", Zurich IEER Working Paper No. 34.
- Fehr, E., S. Gächter and G. Kirchsteiger (1997)**, "Reciprocity as a contract enforcement device - experimental evidence", *Econometrica*, 65(4), pp. 833-860.
- Fehr, E., A. Klein and K.M. Schmidt (2007)**, "Fairness and contract design", *Econometrica*, 75(1), pp. 121-154.
- Fehr, E., G. Kirchsteiger and A. Riedl (1993)**, "Does fairness prevent market clearing? An experimental investigation", *Quarterly Journal of Economics*, 108(2), pp. 437-460.
- Fehr, E. and K.M. Schmidt (1999)**, "A theory of fairness, competition and cooperation", *Quarterly Journal of Economics*, 114(3), pp. 817-868.
- Festinger, L. (1957)**, *A Theory of Cognitive Dissonance*. Stanford University Press, Stanford.
- Figuières, C., D. Masclet and M. Willinger (2012)**, "Vanishing leadership and declining reciprocity in a sequential contribution experiment", *Economic Inquiry*, 50(3), pp. 567-584.
- Fischbacher, U., S. Gächter and E. Fehr (2001)** "Are people conditionally cooperative? Evidence from a public a goods experiment", *Economics Letters* 71(3), pp. 397-404.

- Flood, M.M. (1952)**, "Some experimental games", RAND Research Memorandum, RM-789.
- Forsythe, R., J.L. Horowitz, N.E. Savin and M. Sefton (1994)**, "Fairness in simple bargaining experiments", *Games and Economic Behavior*, 6(3), pp. 347-369.
- Friedman, J. (1971)**, "A noncooperative equilibrium for supergames", *Review of Economic Studies*, 38(1), pp. 1-12.
- Frohlich, N. and J. Oppenheimer (1998)**, "Some consequences of e-mail vs. face-to-face communication in experiment", *Journal of Economic Behavior and Organization*, 35(3), pp. 389-403.
- Fudenberg, D. and E. Maskin (1986)**, "The folk theorem in repeated games with discounting or with incomplete information", *Econometrica*, 54(3), pp. 533-556.
- Gabor, A. and J.T. Mahoney (2010)**, "Chester Barnard and the systems approach to nurturing organizations", University of Illinois, College of Business Working Papers Series, n°10-0102.
- Gächter, S., D. Nosenzo, E. Renner and M. Sefton (2010)**, "Sequential vs. simultaneous contributions to public goods: experimental evidence", *Journal of Public Economics*, 94(7-8), pp. 515-522.
- Geneakoplos, J., D. Pearce and E. Stachetti (1989)**, "Psychological games and sequential rationality", *Games and Economic Behavior*, 1(1), pp. 60-79.
- Gibbons, R. (1998)**, "Incentives in Organizations", *Journal of Economic Perspectives*, 12(4), pp.115-132.
- Gintis, H. (2000)**, "Strong reciprocity and human society", *Journal of Theoretical Biology*, 206(2), pp. 169-179.
- Granovetter, M. (1985)**, "Economic Action and Social Structure: The Problem of Embeddedness", *American Journal of Sociology*, 91(3), pp. 481-510.
- Granovetter, M. (2005)**, "The Impact of Social Structure on Economic Outcomes", *Journal of Economic Perspectives*, 19(1), pp. 33-50.
- Grossman, S.J. and O. Hart (1983)**, "An analysis of the principal-agent problem", *Econometrica*, 51(1), pp. 7-46.
- Güth, W., S. Huck and W. Müller (2001)**, "The relevance of equal splits in ultimatum games", *Games and Economic Behavior*, 37(1), pp. 161-169.

- Güth, W., M.V. Levati, M. Sutter and E. van der Heijden (2007)**, "Leading by example with and without exclusion power in voluntary contribution experiments", *Journal of Public Economics*, 91(5-6), pp. 1023-1042.
- Güth, W., R. Schmittberger and B. Schwarze (1982)**, "An experimental study of ultimatum bargaining", *Journal of Behavior and Organizations*, 3(4), pp. 367-388.
- Güth, W. and R. Tietz (1990)**, "Ultimatum bargaining behavior. A survey and comparison of experimental results", *Journal of Economic Psychology*, 11(3), pp. 417-449.
- Güth, W. and E. van Damme (1998)**, "Information, strategic behavior, and fairness in ultimatum bargaining: an experimental study ", *Journal of Mathematical Psychology*, 42(2-3), pp. 227-247.
- Hamburger, H., M. Guyer and J. Fox (1975)**, "Group size and cooperation", *Journal of Conflict Resolution*, 19(3), pp. 503–531.
- Hart, O. and B. Holmstrom (1987)**, "The Theory of contracts", in T. Bewley (ed.), *Advances in Economic Theory, Fifth World Congress*, Cambridge University Press, Cambridge.
- Hauk, E. (2003)**, "Multiple prisoner's dilemma games with(out) an outside option: An experimental study", *Theory and Decision*, 54(3), pp. 207-229.
- Head, J.G. (1962)**, "Public Goods and Public Policy", *Public Finance*, 17(3), pp. 197-219.
- Hechter, M. and K.D. Opp (2001)**, Introduction, in M. Hechter and K.D. Opp (eds.), *Social Norms*, Sage, New York, pp. xi-xx.
- Hichri, W. and A. Kirman (2007)**, "The emergence of coordination in public good games", *The European Physical Journal B*, 55(2), pp. 149-159.
- Hobbes, T. (1909)**, *Leviathan*, Clarendon Press, Oxford [1651].
- Hodgson, G.M. (2004a)**, "Opportunism is not the only reason why firms exist: why an explanatory emphasis on opportunism may mislead management strategy", *Industrial and Corporate Change*, 13(2), pp. 401-418.
- Hodgson, G.M. (2004b)**, "Reclaiming habits for institutional economics", *Journal of Economic Psychology*, 25(5), pp. 651-660.
- Holmstrom, B. (1979)**, "Moral hazard and observability", *Bell Journal of Economics*, 10(1), pp. 74-91.

- Holmstrom, B. (1982)**, "Moral Hazard in Teams", *Bell Journal of Economics*, 13(2), pp. 324-340.
- Holt, C.A. and S.K. Laury (2008)**, "Theoretical explanations of treatment effects in voluntary contributions experiments", in C.R. Plott and V.L. Smith (eds.), *Handbook of Experimental Economics Results Volume I*, North-Holland, Oxford.
- Horne, C. (2001)**, Sociological Perspectives on the Emergence of Social Norms, in M. Hechter and K.D. Opp (eds.), *Social Norms*, Sage, New York, pp. 3-34.
- Isaac, R. M. and J. Walker (1988a)**, "Group size effects in public goods provision: The voluntary contributions mechanism", *Quarterly Journal of Economics*, 103(1), pp. 179-199.
- Isaac, R. M. and J. Walker (1988b)**, "Communication and free-riding behaviour: The voluntary contribution mechanism", *Economic Inquiry*, 26(4), pp. 585-608.
- Isaac, R. M., J. Walker and S. Thomas (1984)**, "Divergent evidence on free riding: An experimental examination of possible explanations", *Public Choice*, 43(1), pp. 113-149.
- Isaac, R. M., J. M. Walker, and A. Williams (1994)**, "Group size and the voluntary provision of public goods: experimental evidence utilizing large groups", *Journal of Public Economics* 54(1), pp. 1-36.
- Jensen, M.C. and W.H. Meckling (1976)**, "Theory of the Firm: Managerial Behavior, Agency Cost, And Ownership Structure", *Journal of Financial Economics*, 3(4), pp. 305-360.
- Kahneman, D., J.L. Knetsch and R.H. Thaler (1986)**, "Fairness and the assumptions of economics", *Journal of Business*, 59(4), pp. S285-S300.
- Keser C., A. Markstädter, M. Schmidt and C. Schnitzler (2011)**, "Rich man and Lazarus - asymmetric endowments in public-goods experiments", Working Paper, Göttingen University.
- Keser, C., and C. Montmarquette (2004)**, "Voluntary teaming and effort", Cirano Scientific Series: 2004s-49.
- Keser, C., and F. van Winden (2000)**, "Conditional cooperation and voluntary contributions to public goods", *Scandinavian Journal of Economics*, 102(1), pp. 23-39.
- Kinokawa, S., T. Saijo and M. Une (2000)**, "Partial communication in a voluntary-contribution-mechanism experiment", *Pacific Economic Review*, 5(3), pp. 411-428.

- Kollock, P. (1998)**, "Social dilemma: the anatomy of cooperation", *Annual Review of Sociology*, 24, pp. 183-214.
- Kuhlman, D.M. and A.F.J. Marshello (1975)**, "Individual differences in game motivation as moderators of preprogrammed strategy effects in prisoner's dilemma", *Journal of Personality and Social Psychology*, 32(5), pp. 922-931.
- Ledyard, J. O. (1995)**, "Public goods: A survey of experimental research", in J. H. Kagel and A. R. Roth (eds.), *Handbook of Experimental Economics*, Princeton University Press, pp. 111-194.
- Macy M.W. (1993)**, "Social learning and the structure of collective action", in Lawler, E., B. Markovsky, K. Heimer and J. O'Brien, *Advances in Group Processes, Volume 10*, JAI Press, Greenwich.
- Mahoney, J.T. (2005)**, *Economic Foundations of Strategy*, Sage, Thousand Oaks.
- Marwell, G. and R.E. Ames (1979)**, "Experiments on the provision of public goods. I. Resources, interest, group size, and the free-rider problem ", *American Journal of Sociology*, 84(6), pp. 1335-1360.
- Marwell, G. and R.E. Ames (1980)**, "Experiments on the provision of public goods. II. Provision Point, Stakes, Experience, and the free-rider problem ", *American Journal of Sociology*, 85(4), pp. 926-937.
- Masclot, D., C. Noussair, S. Tucker and M.C. Villeval (2003)**, "Monetary and non-monetary punishment in the voluntary contributions mechanism", *American Economic Review*, 93(1), pp. 366-380.
- Maurice, J., A. Rouaix and M. Willinger (2011)**, "Income redistribution and public good provision: an experiment", LAMETA Working Paper 09-12, University of Montpellier.
- McClintock, C.G. and W.B.G. Liebrand (1988)**, "Role of interdependence structure, individual value orientation, and another's strategy in social decision making: a transformational analysis", *Journal of Personality and Social Psychology*, 55(3), pp. 396-409.
- Meckling, W. H. (1976)**, "Values and the Choice of the Model of the Individual in the Social Sciences", *Swiss Journal of Economics and Statistics*, 112(4), pp. 545-560.
- Mill, J. S. (1844)**, *Essays on some Unsettled Questions of Political Economy*, John W. Parker, London [<http://www.archive.org/details/essaysonsomeuns01millgoog>, 2006].

- Nalbantian, H.R. And A. Schotter (1997)**, "Productivity under group incentives: an experimental study", *American Economic Review*, 87(3), pp. 314-341.
- Neal, D.T., W. Wood and J.M. Quinn (2006)**, "Habits - a repeat performance", *Current Directions in Psychological Science*, 15(4), pp. 198-202.
- Nelson, R.R. and S.G. Winter (1982)**, *An Evolutionary Theory of Economic Change*, Harvard University Press, Cambridge.
- Nonaka, I. and H. Takeuchi (1995)**, *The Knowledge-Creating Company*, Oxford University Press, New York.
- Normann, H.-T. and H. Rau (2011)**, "Step-level public goods: experimental evidence", Working Paper, Düsseldorf Institute for Competition Economics.
- Olson, M. (1971)**, *The Logic of Collective Action: Public Goods and the Theory of Groups*, Harvard University Press, Cambridge [1965].
- Ones, U. and L. Putterman (2007)**, "The ecology of collective action: a public goods and sanctions experiment with controlled group formation ", *Journal of Economic Behavior and Organization*, 62(4), pp. 495-521.
- Orbell, J., and R. Dawes (1993)**, "Social welfare, cooperator's advantage, and the option of not playing the game", *American Sociological Review*, 58(6), pp. 787-800.
- Orbell, J., P. Schwartz-Shea and R. Simmons (1984)**, "Do cooperators exit more readily than defectors?", *American Political Science Review*, 78(1), pp. 147-162.
- Orbell, J.M., A.J.C. van de Kragt and R.M. Dawes (1988)**, "Explaining discussion-induced cooperation", *Journal of Personality and Social Psychology*, 54(5), pp. 811-819.
- Ostrom, E. (2005)**, "Policies that crowd out reciprocity and collective action", in H. Gintis, S. Bowles, R. Boyd, & E. Fehr (Eds.), *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*, MIT Press, Cambridge, pp. 253-275.
- Ostrom, E., J. Walker and R. Gardner (1992)**, "Covenants with and without a sword: self-governance is possible", *American Political Science Review*, 86(2), pp. 404-417.
- Page, T., L. Putterman and B. Unel (2005)**, "Voluntary association in public goods experiments: Reciprocity, mimicry and efficiency", *Economic Journal*, 115(506), 1032–1053.
- Pascal, B. (1963)**, *Pensées*, Seuil, Paris [1669] and translation by W. F. Trotter, *Thoughts*, P.F. Collier & Son, New York [1909-1914].

- Polanyi, K. (1957)**, *The Great Transformation: the Political and Economic Origins of our Time*, Beacon Press, Boston.
- Potters, J., M. Sefton and Lise Vesterlund (2007)**, "Leading-by-example and signaling in voluntary contribution games: an experimental study", *Economic Theory*, 33(1), pp. 169-182.
- Prendergast, C. (1999)**, "The provision of incentives in firms", *Journal of Economic Literature*, 37(1), pp. 7-63.
- Rabin, M. (1993)**, "Incorporating fairness into game theory and economics", *American Economic Review*, 83(5), pp. 1281-1302.
- Rapoport, A. and A.M. Chammah (1965)**, *Prisoner's Dilemma: A Study in Conflict and Cooperation*, University of Michigan Press, Ann Arbor.
- Rapoport, A. and D. Eshed-levy (1989)**, "Provision of step-level public goods: effects of greed and fear of being gypped", *Organizational Behavior and Human Decision Processes*, 44(3), pp. 325-344.
- Ravaisson, F. (1838)**, *De l'Habitude*, H. Fournier, Paris [Digitalized by A. Blachair, Académie de Nancy-Metz, available at [www.ac-nancy-metz.fr/enseign/philo/textesph](http://www.ac-nancy-metz.fr/enseign/philo/textesph)].
- Rawls, J. (1971)**, *A theory of Justice*, Harvard University Press, Cambridge.
- Sefton, M., R. Shupp and J.M. Walker (2007)**, "The effect of rewards and sanctions in provision of public goods", *Economic Inquiry*, 45(4), pp. 671-690.
- Shafir, E. and A. Tversky (1992)**, "Thinking through uncertainty: nonconsequential reasoning and choice", *Cognitive Psychology*, 24(4), pp.449-474.
- Simon, H.A. (1941)**, *A Theory of Administrative Decision*, PhD Dissertation, Chicago.
- Simon, H.A. (1997)**, *Administrative Behavior: a Study of Decision-making Processes in Administrative Organizations*, The Free Press, New York [1947].
- Smith, A. (1902)**, *An Inquiry into the Nature and Causes of the Wealth of Nations*, P.F. Collier & Son, New York [<http://www.archive.org/details/wealthofnations00smituoft>, 2008] [1776].
- Stahl, I. (1972)**, *Bargaining Theory*, Economic Research Institute, Stockholm.
- Summer, W.G. (1906)**, *Folkways: a Study of the Sociological Importance of Usages, Manners, Customs, Mores, and Morals*, Ginn and Co., Boston.



- Tajfel, H. and J. Turner (1979)**, "An integrative theory of intergroup conflict", in W. Austin and S. Worchel (eds.), *The Social Psychology of Intergroup Relations*, Brooks/Cole, Monterrey, pp. 33-47.
- Tan, J.H.W. And F. Bolle (2007)**, "Team competition and the public good game", *Economic Letters*, 96(1), pp. 133-139.
- Teece, D.J., R. Rumelt, G. Dosi and S.G. Winter (1994)**, "Understanding Corporate Coherence: Theory and Evidence", *Journal of Economic Behavior and Organization*, 23(1), pp. 1-30.
- Trivers, R.L. (1971)**, "The evolution of reciprocal altruism", *Quarterly Review of Biology*, 46(1), pp. 35-57.
- Tucker, A.W. (1950)**, "A two-person dilemma", Mimeo, Stanford University.
- van de Kragt, A.J.C., J.M. Orbell and R.M. Dawes (1983)**, "The minimal contributing set as a solution to public goods problems", *American Political Science Review*, 77(1), pp. 112-122.
- van Dijk, E. and H. Wilke (1994)**, "Asymmetry of wealth and public good provision ", *Social Psychology Quarterly*, 57(4), pp. 352-359.
- van Lange, P.A.M., W.B.G. Liebrand, D.M. Messick and H.A.M. Wilke (1992)**, "Social dilemma: the state of the art. Introduction and literature review", in W.B.G. Liebrand, D.M. Messick and H.A.M. Wilke (eds), *Social Dilemmas: Theoretical Issues and Research Findings*, Pergamon, Oxford.
- Veblen, T. (2003)**, *The Theory of the Leisure Class*, Pennsylvania State University, Hazleton [1899].
- Williamson, O.E. (1975)**, *Markets and Hierarchies: Analysis and Antitrust Implications*, The Free Press, New York.
- Williamson, O.E. (1985)**, *The Economic Institutions of Capitalism*, The Free Press, New York.
- Williamson, O.E. (1991)**, The Logic of Economic Organization, in O.E. Williamson and S.G. Winter (eds.), *The Nature of The Firm: Origins, Evolution and Development*, Oxford University Press, New York, pp. 90-116.
- Wood, W. and D.T. Neal (2007)**, "A New Look at Habits and the Habit-Goal Interface", *Psychological Review*, 114(4), pp. 843-863.

# CONTENTS

SHORT CONTENTS.....	11
GENERAL INTRODUCTION.....	13
PART ONE.....	15
CHAPTER I	
Economic theories of collective production and cooperation.....	17
1.Introduction.....	18
2.Enforcing cooperation: a matter of incentives.....	19
2.1.A familiar picture: the selfish agent.....	19
2.2.Property rights.....	20
2.3.Agency theory.....	20
2.4.Multiple agents: teams.....	21
2.5.Transaction cost economics (TCE).....	24
3.Mancur Olson and The Logic of Collective Action.....	25
3.1.Group size and individual weights.....	26
3.2.Inclusive vs. exclusive groups.....	27
3.3.Social incentives and the mobilization of latent groups.....	27
4.Barnard, Simon, Cyert, March and the behavioral theory of the firm.....	28
4.1.Barnard: moral behavior and informal organization.....	28
4.2.Simon: bounded rationality and limited attention.....	30
4.3.Cyert and March: subcoalitions and bargaining.....	32
4.4.The "behavioral" agent.....	32
5.Learning organizations: knowledge, routines and communities.....	33
6.The game theoretical analysis of social dilemma.....	35
6.1.Nash vs. Pareto optimality.....	35
6.2.The evolutionary game theory approach: reciprocity and punishment.....	35
6.3.Taking the utility function one step beyond: fairness and equity.....	36
7.Looking forward: Norms, institutions, habits and motivation.....	38
7.1.Social norms and social preferences.....	38
7.2.Institutions and habits.....	40
7.3.Intrinsic motivation vs. extrinsic incentives?.....	41
8.Conclusion: irreconcilable views?.....	43
CHAPTER II	
Illustrating the double edge of cooperation.....	45

1.Introduction.....46

2.Symmetric two-person interactions: prisoner's dilemma experiments.....48

    2.1.Stylized facts: cooperation happens.....48

    2.2.The sequential prisoner's dilemma game.....49

    2.3.Opening the prison: voluntary participation.....50

3.Symmetric, n-person interactions: public good game and n-prisoner's dilemma experiments.....51

    3.1.A misleading stylized fact: positive, decaying contributions.....52

    3.2.Parameters: Group size and multiplier value.....53

    3.3.Communication.....54

    3.4.Commitment, promises or 'covenants without a sword'.....57

    3.5.Endowment value and symmetry.....58

    3.6.Provision points and thresholds.....60

        Critical mass.....61

        Provision point experiments.....61

    3.7.Sequential contributions and leadership.....63

    3.8.Refinements of the mechanism 1 : sanctions and rewards.....64

        A different type of sanctions: exclusion.....66

    3.9.Refinements of the mechanism 2 : manipulating matching patterns.....67

        Exogenous group formation.....67

        Endogenous group formation.....68

    3.10.Group incentives and team competition.....69

4.Asymmetric two-person interactions: principal-agent settings, the 'gift-exchange game' and the ultimatum/dictator games.....70

    4.1.Incentives, control and trust.....71

        Introducing incentives.....72

    4.2.Contract choice.....74

    4.3.Monitoring intensity: extrinsic incentives and intrinsic motivation.....76

    4.4.Fairness and equity: ultimatum and dictator games.....77

        Ultimatum game experiments.....77

        Dictator game experiments.....78

5.Asymmetric n-person interactions: three-player distribution games.....79

6.Conclusions.....81

    6.1.Smart self-interest: far-sightedness, reciprocity and conditional cooperation.....81

6.2. Broadened self-interest: conformity, equity and "altruism".....	82
6.3. Indecisiveness of collective production situations.....	82
6.4. Three steps further.....	83
PART TWO.....	85
CHAPTER III	
Voluntary participation in a collective-good game.....	87
1. Introduction.....	88
2. Related literature on voluntary participation.....	89
3. Experimental settings and treatments .....	91
3.1. The baseline (B) treatment.....	92
Theoretical predictions and hypothesis.....	93
3.2. The voluntary participation ( VP) treatments.....	94
Theoretical predictions and hypothesis.....	96
4. Results and discussion .....	96
4.1. Choices heterogeneity: neither Nash nor the optimal social outcome .....	97
4.2. Observed average behavior as a strong attractor.....	97
4.3. Treatments VP vs. treatment B.....	101
Average contribution levels.....	101
Contribution rates over time.....	102
Group dynamics.....	104
Individual behavior: the exit option as a (non-)credible threat?.....	106
5. Comments .....	107
6. Conclusion.....	109
CHAPTER IV	
Alternative designs for cooperation .....	111
1. Introduction.....	112
2. The experiment.....	115
2.1. The principal-agent games.....	115
2.2. Game-theoretical solution.....	116
2.3. Experimental settings.....	117
3. Results.....	118
3.1. Contract choices : Partner Treatment (PT) vs Stranger Treatment (ST).....	118
Shares of IC vs TC.....	119
Contract switches: the relative stability of contract choices.....	122

3.2.Contract efficiency: Partner Treatment (PT) vs Stranger Treatment (ST).....	123
3.3.Contract efficiency: Incentive Contract vs Trust Contract.....	124
3.4.Reciprocity: the determinants of effort levels.....	126
4.Discussion.....	128
5.Conclusion.....	129
CHAPTER V	
Disentangling distributional motives.....	131
1.Introduction.....	132
2.Related literature.....	132
2.1.Ultimatum and dictator games.....	132
2.2.Efficiency, maximin and the two models of inequality aversion .....	133
2.3.A renewed case for inequality aversion?.....	135
3.A simple distribution game.....	137
3.1.Refining the settings.....	137
3.2.Meeting John Rawls.....	138
3.3.Maximin vs. 'efficiency'.....	140
3.4.Equity vs. 'selfishness'.....	142
4.Results.....	143
4.1.Choosing a distribution: points victory for equity.....	143
4.2.The hidden effects of risk and uncertainty: intra-player analysis.....	145
5.Discussion.....	146
6.Conclusion.....	148
GENERAL CONCLUSION.....	151
APPENDICES.....	155
Appendix A: Instructions to subjects - Chapter 3.....	155
Appendix B: Instructions to subjects - Chapter 4.....	159
Appendix C: Instructions to subjects - Chapter 5.....	164
REFERENCES.....	167



**COLLECTIVE PRODUCTION PROCESSES,  
COOPERATION AND INCENTIVES  
EXPERIMENTAL EXPLORATIONS**

## Résumé

L'étude des processus de création de connaissances souligne la complexité des interactions individuelles au sein des organisations productives. Cette complexité est telle que les théories de l'entreprise basées sur les incitations, focalisées sur les problèmes de traitement de l'information, peuvent ignorer une part substantielle des facteurs de décision individuels intervenant dans le contexte de l'apprentissage organisationnel, et plus largement dans les processus de production collectifs. Nous utilisons dans cette thèse la méthode expérimentale pour étudier les déterminants de la coopération, afin d'affiner les hypothèses comportementales sur lesquelles sont basées les théories économiques de la production collective. Nous montrons que les deux visions de la coopération portées par les théories de l'entreprise - un comportement devant être extrait d'intérêts divergents et une propriété émergente découlant des interactions sociales entre agents - sont étayées par les résultats expérimentaux. Par conséquent, nous concluons que les deux approches devraient être retenues et éventuellement intégrées dans un cadre d'analyse plus large.

Mots-clés: Coopération, Théorie de l'entreprise, Théorie des organisations, Économie expérimentale

## Abstract

The study of knowledge creation processes has pointed to the complexity of individual interactions within productive organizations. This complexity appears to be such that incentive-based theories of the firm, which focus on information processing issues, may fail to grasp a substantial part of the individual decision-making involved in the context of organizational learning, and more broadly in collective production processes. In this thesis we use experimental methods to study the determinants of cooperation, in order to refine the behavioral assumptions on which economic theories of collective production are based. We show that the two visions of cooperation embodied in competing theories of the firm - a behavior to be elicited from diverging interests and an emergent property stemming from social interactions among agents - find support from the laboratory experiments. Accordingly, we conclude that both approaches should be upheld and possibly combined in a broader, integrative, analytical framework.

Keywords: Cooperation, Theory of the firm, Organization theory, Experimental economics