



## UvA-DARE (Digital Academic Repository)

### Keeping up appearances: Experiments on cooperation in social dilemmas

van den Broek, E.M.F.

**Publication date**  
2014

[Link to publication](#)

#### **Citation for published version (APA):**

van den Broek, E. M. F. (2014). *Keeping up appearances: Experiments on cooperation in social dilemmas*. Rozenberg.

#### **General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### **Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# Introduction

We cooperate. To be precise: people reduce their own chances of survival and reproduction to augment those of others, daily and at an astonishing scale; among family, strangers and enemies, whether we are foraging, negotiating or waging war. Why are we willing to help out others at a cost to ourselves? One reason is that collectives can accomplish more than individuals. That applies to cells, ants, monkeys, persons and firms. But our taste for altruism goes beyond rational calculations. We behave altruistically even when our kindness cannot be observed, let alone returned by anyone else.

This poses a puzzle to economists and biologists alike. Both disciplines have dwelled for centuries on (respectively) welfare or fitness sacrificing behavior. Let me discuss two excerpts that illustrate the stance of the foremost thinkers from each side of the sciences.

*“In the first place, as the reasoning powers and foresight of the members became improved, each man would soon learn that if he aided his fellow-men, he would commonly receive aid in return.”*

In this quote, Charles Darwin (1871) clearly formulates a rationale for human altruism. He approached human social behavior as any other animal’s social behavior: as a means to survive and proliferate. This strategic account of our behavior foregoes, however, the pleasure that we seem to derive from doing good. Such an ‘irrational’ warm glow is mentioned by the 18<sup>th</sup> century economist Adam Smith (1759):

*“How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it, except the pleasure of seeing it.”*

Darwin’s statement can be viewed as the strategic, fitness-enhancing account of altruism, whereas Smith’s statement addresses the proximate psychological mechanism by which evolution ‘implemented’ altruistic behavior in humans: we are more inclined to behave

altruistically if the delay until the reward is short. Precisely this psychological mechanism makes us vulnerable for exploitation.

In this thesis I address the issue of human cooperation. I study how the institutional environment steers behavior in situations in which an individual's interest is not in line with the general interest of the population at large. I use laboratory experiments to assess the effect of information about (others') previous behavior on cooperative decisions. In Chapter 2, I investigate the effects of reputations in games played by two players who meet repeatedly, or, alternatively, have some information about previous behavior of their partner towards others. In Chapter 3 and 4 I study games in which groups of people interact and obtain knowledge about each other or about other groups.

The main instrument for analyzing situations in which decision makers face strategic dilemmas is game theory. This was systematically developed by Von Neumann and Morgenstern (1944) as a set of mathematical tools to model multi-player situations. It does so by defining a set of players, an action space for all players and their preferences over outcomes. A prominent solution concept for such a game, the Nash equilibrium, is defined as a set of actions for which no player has an incentive to deviate, given that he knows that others play the equilibrium action, too. Game theory is grounded in a rational actor model, which is ubiquitous in economics for two reasons. First, any evolved life form has been adapted to an environment in such a way that it exploits regularities for self-preservation, so we may assume that whatever an individual does, it is consistent with optimizing some kind of preference function (Maynard Smith, 1958); second, although the assumptions about the behavior of individuals may be far-fetched, they produce testable predictions at the aggregate level. This is considered to be a good thing, and many such predictions (for example market equilibrium in a competitive environment) find empirical support. Nevertheless, these predictions hinge on strong assumptions, such as complete information, stable preferences and unlimited computing power of the individuals. Each of these assumptions has been challenged and refined (Simon, 1976; Selten, 1990; Tversky and Kahneman, 1974; Gigerenzer and Selten, 2001).

These extensions to the rational actor model have emerged partly from the surge of experimental economics around 1980 as a new field of science. The first market experiments in the 1960s were introduced as a way to test the external validity of economic models (Davis and Holt, 1993). The incongruities between game theoretical predictions and empirical findings ('anomalies') slowly motivated researchers, journals and institutions towards incorporating psychological evidence into economic theories. Over the past twenty years, some of the self-regarding agents have been pimped up with a more descriptive preference function that incorporates the well-being of others (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), fairness considerations (Andreoni and Miller, 2002), intentions (Rabin, 1993), or beliefs (Dufwenberg and Kirchsteiger 2004).

## ***1.1 Methodology***

Controlled laboratory experiments allow us to study the decisions of participants who find themselves on the brink of our disposition to do good and to be exploited by others. The situations we create for them are abstract translations of everyday situations: the decision to trust a stranger, whether to help someone with a good reputation or whether to join a group. In contrast to the outside world, the laboratory offers us an environment in which we can control the information each subject has. Stripping down the decision problem to its incentive structure allows us to focus on the strategic considerations of participants in the experimental setting. Experimental economists foster a treasury of such set-ups or games. By introducing isolated features of real life we can show their influence on behavior. Below I provide a short overview of the games that are central to this thesis, their theoretical predictions and the stylized facts of the empirical data.

### *Prisoner's dilemma*

Perhaps the most famous game is the Prisoners' dilemma. Consider a situation in which two suspects of a crime face the decision to either cooperate (remain silent with respect to their involvement) or defect (witness against the other). Both are better off when neither of them confesses, since then they receive a short sentence of one year; but if only one of the two confesses she will be set free and the other punished harshly with ten years. If both confess, however, they both face a long sentence: eight years (see table 1.1; note that payoffs –the numbers in the table– are negative because it is assumed that people do not like to be in prison). This situation is characterized by uncertainty about the other's choice, strategic reasoning and a tension between own and other's wellbeing. Note that irrespective of what one prisoner does, her partner in crime is better off defecting by confessing the crime.

**Table 1.1 Prisoner's Dilemma**

$\text{♀} \backslash \text{♂}$	Don't confess	Confess
Don't confess	-1, -1	-10, 0
Confess	0, -10	-8, -8

*Notes.* Payoff table for the two players for each of their actions. The first number in each cell is the payoff for player ♀, the second number is the payoff for player ♂.

This game does not feature in the thesis, but is presented here because it shows the descriptive limits of game theory: in laboratory situations and in real life, people behave much more cooperatively than predicted, even if the game is played only once- see for instance the final part of the UK television show “Golden Balls” called “Split or Steal”.<sup>1</sup>

#### *Helping game*

The helping game is the first half of an alternating variation of the Prisoners' dilemma described above. A player in the role of donor faces the decision to allocate a benefit (e.g., €2,50) to a fellow player – the recipient - at a certain cost (e.g., €1,50), which is lower than the benefit for the recipient. If the donor decides not to help the recipient, their payoffs do not change (see table 1.2 for the payoff matrix).

---

<sup>1</sup> There is a Dutch equivalent of this show known as “Deelt ‘ie het of deelt ‘ie het niet” that has been analyzed by Belot et al. (2010).

**Table 1.2 Helping game**

Donor \ recipient	
Defect	0,0
Help	€ -1,50, € 2,50

*Notes.* Payoff for the donor and recipient when they defect or help.

When two players of this game face each other repeatedly up to infinity, defection is not the best strategy (Trivers, 1971). If the chance of encountering a partner again is sufficiently high, direct reciprocal strategies such as Tit-for-Tat thrive (Axelrod and Hamilton, 1981). If the game is repeated among a group of players, a player may face the donor of some other player and have to rely on reputation. In that case, indirect reciprocity, or “I help you and somebody else helps me” is a feasible strategy (Alexander 1987), as can be observed in real life and in the laboratory (Seinen and Schram, 2006). A simple representation of previous interactions has been formalized in the seminal paper by Nowak and Sigmund (1998), who showed that *image scoring* can sustain cooperation. In Chapter 2, we use this game to study situations in which people can choose between information on direct and indirect reciprocity.

### *Public goods game*

A public goods game captures a social dilemma in which the individual interest of group members conflicts with the collective interest. It is a generalized multi-person version of the Prisoners’ dilemma. A typical public goods experiment consists of a number of rounds. Each round the participants receive a certain amount of money, for instance 1 euro. Independently and simultaneously they decide how much of this to contribute to a public good, and how much to keep in their personal account. The total of contributions to the public good is multiplied by a given factor and distributed evenly among all players<sup>2</sup>. Hunting, air pollution, and overfishing are all group activities that boil down to a public goods game and share the problem of how to keep up cooperation instead of unraveling into free riding singularities. Laboratory studies of such games show a surprising amount of collaboration in early rounds,

---

<sup>2</sup> Note that if the number of players exceeds the multiplication factor, it is most profitable for each individual player if she contributes nothing.

but as the dilemma is repeated, members' contributions to the public good decline. Several explanations for this stylized fact exist. Some account for it by learning effects or by interaction between different types (Andreoni 1988; Kurzban and Houser, 2005); others extend the game by providing players with a longer time horizon or allowing them to communicate (Gächter et al., 2008).

### *Punishment*

One of the enforcement mechanisms we employ in real life is punishment. People punish free riders and norm violators, even at a cost to themselves (Ostrom et al., 1992). Recent laboratory evidence shows that the decrease in cooperation in a public goods game can be prevented if the game is extended with a stage during which players are allowed to punish other players at a cost to themselves. This is because such punishment is mostly directed towards non-cooperators who in response change their behavior in subsequent rounds of the game (Fehr and Gächter, 2002). The exertion of punishment is widespread throughout different cultures (Henrich et al., 2006) and has been observed also when players meet only once (Fehr and Fischbacher, 2004), and the effects extend even to non-monetary punishment (Masclot et al., 2003). Therefore costly punishment has been put forward as an explanation for the persistence of cooperation in human societies (Nowak, 2006).

Together, experimental evidence from these games and recent extensions of individual preference functions have succeeded in explaining a larger part of the richness of empirical observations of cooperation and altruism. Still, there remains a gap between evolutionary and economic theories on cooperation and our behavior in firms, in eBay transactions and on the street. Although in economic experiments participants always face real monetary rewards, incorporating reputation in a realistic way in laboratory settings is a stretch (Schram, 2005; Hagen and Hammerstein, 2006). By studying the effect of information about previous behavior, this thesis aims to narrow the gap between life outside and inside the laboratory in situations in which an individual's interest is not in line with the general interest.



## *1.2 Chapter Guide*

We will consider two aspects of the influence of information on cooperation. In Chapter 2, we look at reputation in games played by two players who meet repeatedly, or, alternatively, have some information about previous behavior of their partner. In Chapters 3 and 4 we study games in which groups of people interact. In each of these three Chapters specific game-theoretical predictions are tested by an experiment on human subjects. A general conclusion based on all chapters is provided in Chapter 5. What follows is an overview of the research questions to be answered in the three chapters describing experiments.

In Chapter 2 we study repeated interactions amongst pairs drawn from a group of individuals. People have access to information about their own previous experience with the partner, but also about others' experience with her. In such situations, cooperation can be sustained through direct and indirect reciprocity. This is a novel environment, because previous studies of indirect reciprocity focus on groups in which direct experience is excluded. This study examines experimentally which kind of information people prefer to use, when both types of information are available, yet costly. Simulations show that the use of reputations (i.e., information about experiences of third parties) can have a selective advantage in a population of agents only using information on direct interactions. In our experiments, we find that people request both kinds of information evenly, but assign different weights to them when deciding whether or not to reciprocate. A decrease in the reliability of the indirect information does not affect the cooperation level, but increases the demand for direct information.

Next, we turn to social dilemmas in groups (Chapter 3). Individuals are not simply born in social groups but can choose with whom they interact. We consider the relevance of this for voluntary public good provision in groups, noting that in many environments people choose with whom they prefer to cooperate, and may exclude some others altogether. Previous studies on endogenous group formation in the laboratory showed that voting and rank-based mechanisms can sustain cooperation. We present a laboratory environment with fixed group size, where we allow both for ranking and exclusion of partners based on information from their previous choices. We find that endogenous group formation with exclusion creates high

cooperation and efficiency levels. In combination with punishment, exclusion yields even higher contributions.

In Chapter 4 we investigate the influence of another group on the behavior of a group member in a social dilemma. Group competition has been put forward as explanation for sustained cooperation. Yet punishment costs are detrimental to a group's relative success in such competition with other groups. We investigate the dynamics of altruistic punishment in combination with intergroup competition by conducting a series of experimental public goods games, in which we systematically vary the possibility for social comparison between groups, group competition, and punishment. Our results indicate that the mere presence of another group is sufficient to induce high cooperation and punishment levels. Group competition attenuates punishing behavior. Despite the lower level of exerted punishment, contribution levels remain equally high, however.

Finally, Chapter 5 concludes and discusses the findings from earlier chapters before suggesting open questions and directions for future work. Overall, this thesis aims to extend our knowledge about reputation by zooming in on the sources of reputation and by broadening the perspective to partner choice and intergroup observation. In the conclusion we expand on the insights gleaned from these experiments.