

# The Puzzle of Social Preferences

by

Jordi Brandts\* and Enrique Fatas\*\*

October 2010

## Abstract

We present a brief overview of the experimental economics literature on social preferences. In numerous experiments, economically incentivized subjects are willing to sacrifice part of their material earnings to compensate the kind behavior of others, or will be willing to reciprocate at a non-negligible cost, or even pay a positive price for punishing the behavior of selfish individuals. All these actions are labeled as *social* in economics because there is no apparent way to reconcile them with any reasonable form of pure self-interest. We focus on social dilemma games and want to communicate two main messages. First, research in experimental economics has produced abundant evidence that illustrates the social components of people's preferences. Second, social sanctions of different types play an important role in facilitating cooperative behavior.

\*Universitat Autònoma de Barcelona and Instituto de Análisis Económico (CSIC)

\*\*Universitat de València

## 1-Introduction

In a well-cited paper Fehr and Fischbacher (2002) call economics the *dismal science* because “economists routinely make worst-case assumptions regarding people’s motives” (page C1). The still dominant logic in economics could not be simpler: individuals are exclusively motivated by their material self-interest. The power of this assumption comes from its very simplicity: as a mostly normative assumption, it does not consider any other behavioral regularity coming from real life or experimental tests of human behavior in economic contexts. In sharp contrast with the interest of economists in the diversity of economic decisions (e.g. in consumption), it simplifies human behavior assuming that subjects are homogeneous in their (null) social motivations.

What we know from the use of systematic experimentation with human subjects in economically incentivized experiments is that participants very frequently choose actions that do not maximize their own monetary payoffs. These decisions become systematic when those actions affect others’ payoffs in very different domains: from market behavior to consumer decisions, from tax compliance to employee reactions to changes in compensation schemes and organizational practices. The experimental economics literature calls these ‘other-regarding’ preferences *social preferences* (see Charness and Rabin, 2002).

In this paper we present a brief overview of some experimental evidence of economic behavior in social dilemmas, at odds with the idea of self-interest. In these experiments, economically incentivized subjects are willing to sacrifice part of their material earnings to compensate the kind behavior of others, or will be willing to reciprocate at a non-negligible cost, or even pay a positive price for punishing the behavior of selfish individuals. All these actions are labeled as *social* in economics because there is no apparent way to reconcile them with any reasonable form of pure self-interest.

Our paper only covers a rather small selection of studies in the area, which allows us to make our points. We focus on social dilemma games and want to communicate two main messages. First, research in experimental economics has produced abundant evidence that illustrates the social components of people’s

preferences. Second, social sanctions of different types play an important role in facilitating cooperative behavior.

In the specific domain of social dilemmas - a literature covered in detail by Chaudhuri (forthcoming) - social behavior is typically associated with the existence of some form of conditional cooperation. A significant fraction of subjects are typically reluctant to free ride on the behavior of others, and are willing to cooperate as much as the other group members do. In social dilemmas like linear public goods games, contributions are not purely altruistic, in the sense that they are not unconditional, and depend on the behavior of others. However, this conditional cooperation mechanism does not suffice to guarantee a socially optimal outcome. The existence of a significant proportion of free riders, and the same logic of conditional cooperation, generates an empirical regularity. Contributions typically decline over time.

The introduction of sanctioning mechanisms solves, in most cases, this cooperation problem. Starting with Yamagishi (1986), different forms of sanctioning mechanisms have been tested in the lab. A typical punishment mechanism consists of an additional decision stage. Subjects may contribute (or not) to the public good and sanction free riders (or any other member in their group). The good news is that the public good provision largely benefits from the possibility of sanctioning defectors.

The bad news is that sanctioning is not always optimal, from a collective point of view, when you consider the punishment costs (for both the punisher and the punished). Moreover, sometimes punishment has no effects at all on the contributions of subjects.

## **2-Social preferences in Economics**

An enormous amount of research in experimental economics has been devoted to the analysis of non-selfish preferences. The capital reference regarding experimental research is Kagel and Roth (1995), in which Ledyard (1995) systematically reviews the experimental literature on public goods, at that period

of time. Kagel and Cooper (2010) update this literature, although they intentionally dismiss the literature on public goods games, covered by Vesterlund (2010)<sup>1</sup>.

In economics the idea of social preferences can be easily explained with the help of simple bargaining games, like the ultimatum game. In such a setting, a Proposer submits an offer about how to divide a fixed pie to a Responder. The Responder may accept the division or reject it, and both of them get nothing. In the unique sub-game perfect equilibrium of the game, assuming both subjects are rational and selfish<sup>2</sup> the Proposer offers a negligible proportion of the pie to the Responder, and the Responder accepts it. In sharp contrast with this normative division, participants in experiments acting as Proposers offer around 40% of the pie, while Responders reject typically reject any offer under 30%. These results are robust enough to resist extreme tests in different cultural environments and for relatively large amounts of money.<sup>3</sup>

Forsythe et al. (1994) were probably the first to disentangle strategic and social motivations in experimental bargaining. They compared an ultimatum game with an alternative game in which the Responder has no power to reject the offer (so, the Proposer becomes a Dictator). As there is no strategic interaction between players it is impossible to identify any equilibrium solution of the game. The unique rational decision to be made by a rational selfish Dictator is to give nothing to the Responder. Subjects typically offer less than in the Ultimatum game, but most offers are still positive.

The Dictator game achieved a clear objective. It established that positive offers by Proposers in the Ultimatum game could not be only driven by strategic, purely rational concerns. In a similar way, it is not obvious whether Responders' rejections of positive offers was simply due to an intrinsic disutility associated with an unfair distribution. A rejection could also be associated with an intentional action: punish the Proposer. Blount (1995) analyzed this possibility with two

---

<sup>1</sup> Fehr and Schmidt (2006) and Camerer (2003) in the lab and Rotemberg (2006) in the field are recent alternative reviews of social preferences. Chaudhuri (2010) reviews the literature in public goods games.

<sup>2</sup> Common knowledge of selfish rationality is strictly required.

<sup>3</sup> Heinrich et al. (2001) run these experiments in 15 different *small societies* (with a limited contact with the western world) and Slonim and Roth (1998) and Cameron (1999) found similar results in Indonesia and Slovakia with much larger pies, ranging from one month salary to one fourth of the average annual income per capita.

identical Ultimatum Games in which Proposers could be economically incentivized Proposers (as in Güth et al., 1982), an external observer (with no specific reward) or robots making random choices. Decisions in the two treatments with human Proposers were remarkably similar, and significantly different from those in the robot treatment.

Some papers have tried to model, from a theoretically point of view, social preferences. The most influential papers are probably Bolton and Ockenfels (2000), Fehr and Schmidt (1999) and Charness and Rabin (2002). The first two papers focus on distributional concerns and show how other regarding preferences over income inequality could explain a large number of experimental outcomes, usually in small group bargaining type environments, which the “standard” economic model of strictly selfish preferences failed to organize. In contrast, the same preferences, under different institutions (e.g., competitive markets) produced the standard results.

Distributional models such as Fehr and Schmidt (1999) represent social preferences as “inequity aversion”. Subjects experience an individually variable disutility whenever they receive a payoff different from the payoff received by others. The disutility is typically smaller when they are better than the others (*advantageous inequity*) than when they are worse than the others (*disadvantageous inequity*). Bolton and Ockenfels (2000) incorporates relative payoffs in a non-linear utility function.

Charness and Rabin (2002) tried to disentangle income inequality from intentionality and reciprocity in an intentions based model incorporating participants’ beliefs about other group members’ actions. A similar insight is obtained in Dufwenberg and Kirchsteiger (2004), Cox, Friedman and Gjerstad (2007), Cox, Friedman and Sadiraj (2008) and Falk and Fischbacher (2006). Rabin (1993) is probably the first paper to highlight the relevance of beliefs (first and second order beliefs) about the strategies of the other players to evaluate their intentions and account for other regarding preferences in the utility function.

### **3-Social preferences in social dilemmas**

Bargaining games are zero-sum games with no tension between individual decisions and collective welfare. Some of the most interesting results about social interaction come from a very different family of games in which social welfare (the sum of individual earnings) depends upon the individual decisions of participants. Social psychologists call these games *social dilemmas*, while economists tend to identify social dilemmas with *public goods* provided through voluntary contributions. We typically label this sort of experiments as *Voluntary Contributions Mechanism* (or VCM).

A typical VCM experiment consists of a series of experimental subjects being randomly seated in a laboratory in independent separate groups of a certain size  $n$ . Each participant is given a certain and limited endowment per round, divided in a tokens or Experimental Currency Units (ECU). Within each group, each participant must make a simple allocation task between a private account and a group account. The private account has no additional returns, other than the security of keeping the ECU privately allocated. Each ECU allocated to the group account is multiplied by a certain parameter (strictly greater than 1, strictly smaller than  $n$ ) and equally shared by all  $n$  group members, regardless of their individual contributions to the public good.

The group externality or social return of the public good generates a potential collective benefit (as it is strictly larger than 1). The private benefit from the public good, however, is always smaller than the private one (as the social return is strictly smaller than  $n$ ). Moreover, there is strategic uncertainty when investing privately, but the benefits coming from the public goods eventually depend in every individual decision.

All decisions are made privately, anonymously and simultaneously. Subjects get some information about past decisions (from average to full information about individual decisions) and the game is repeated several times (with or without random re-matching of groups after each round).

The naïve game-theoretic prediction, based in the existence of rational selfish players who know they play game with identical counterparts, is that no one will ever contribute to the public good. Free riding behavior will be the homogeneous behavioral pattern. In other words, for any positive contribution to the public

good, not contributing is the dominant strategy in the sense that there is no other best response of the decision of the others, no matter what the others do. This public goods problem is a *social dilemma* because the optimal allocation for the group would be to fully invest in the public good.

Examination of the data reveals that subjects do contribute something in almost every VCM experiment. However, contributions start around the 40% to 60% of the endowment and decline over time. Individual behavior is extremely dispersed, with almost half of the subjects getting very close to the free riding prediction, and other half persistently choosing a positive contribution to the public good (not inconsistent with the social preferences hypothesis based on the existence of other regarding preferences). Note that positive contribution may be the consequences of mistakes, random choices or a huge variety of social preferences.

The classical survey by Ledyard (1995) identifies three empirically proven factors that enhance cooperation in the laboratory: communication, threshold points and economic incentives (e.g. a higher social return). The role of economic incentives is maybe not surprising. The larger the return of the public good, all other things equal, the more contribution is observed.

The role of the other two mechanisms is probably more interesting. The introduction of a provision point totally transforms the theoretical prediction of the game. Even for rational selfish players not contributing is not always the best response to the actions of the others. If my contribution is critically needed to make the whole group achieve the goal of reaching the threshold, I may be better off contributing (getting the benefits of everybody else's contribution) than not contributing at all (and getting no public good in return). In other words, there is no way to free ride on the contributions of others if the threshold is not reached (and the public good is not provided). The drawback of this contribution increase is that the provision of the public good is not systematically larger. Many individual contributions get lost, as they don't generate more public good.

Communication generates an increase in contributions even when it is implemented in an anonymous environment and it is non-binding (e.g. subjects may or may not make a decision in line with their promises when *cheap* talking), as in Dawes, McTavish and Shaklee (1977) and Isaac and Walker (1988).

The impact of communication and thresholds are interesting when considering the large heterogeneity in social preferences. Players differ not only in their social preferences (the intensity of their other regarding preferences), but also in their beliefs about the distribution of social preferences among their teammates. Probably one of the clearest regularities found in experimental public goods games is the existence of a significant proportion of conditional cooperators. Contribution is positively correlated with their beliefs about the contributions to be made by their group members.

This heterogeneity was originally attributed to different causes by experimentalists. It was partially explained by subjects' confusion, as in Andreoni (1995), or as a consequence of a noisy and slow learning process of the strategic interaction as in Andreoni (1988), Palfrey and Prisbrey (1997) or Anderson, Goeree and Holt (1998). More recently, heterogeneity has been used to explain the decline in contributions observed in many experiments. Fischbacher, Gächter and Fehr (2001) find that 50% of their subject pool follows a behavioral pattern consistent with the idea of conditional cooperation, even when they typically contribute on average less than the last average contribution. This "self-serving" bias (matching everybody else's contribution at an individually profitable discount) generates the decline in contributions.

The intuition is simple. Any mix of conditional co-operators and free riders will make conditional co-operators to adjust their individual contributions to the last observed average contribution. Heterogeneity makes conditional cooperation to adjust downwards their contributions given the presence of free riders. Keser and van Winden (2000) find that 80% of subjects use average group contributions as a reference point to adjust their individual decisions. Sonnemans, Schram and Offerman (1999) use a clean within and between subjects in which subjects leave their group following a previously known pattern. Subjects contribute very little immediately before they leave a group but are strongly conditional cooperators while remaining in the same team.

Kurzban and Houser (2005) first establish the type of their subjects in a sequence of public goods games. The distribution of types is very similar to other papers mentioned above: 63% are conditional cooperators, 20% are free-riders and 13%



are unconditional cooperators. These patterns are stable across time and repetitions of the game with different opponents.

Fischbacher and Gächter (2009, 2010) use both incentivized experiments and questionnaires to elicit preferences and beliefs about other subjects' average contribution. Experimental decisions and questionnaire answers are roughly consistent: 55% of participants are conditional cooperators and 23% of them selfish free-riders. Eliciting beliefs after making decisions on contributions does not generate any sizeable effect on decisions. Guillem et al (2010) show how beliefs can be optimally manipulated in the lab.

Burlando and Guala (2005) check that this distribution is robust to the use of strategy method, and the decomposed game technique (see Offerman et al. , 1996). Kocher, Cherry, Kroll, Netzer and Sutter (2008), Hermann and Thoni (2008) replicate these results using a heterogeneous subject pool, running the experiments in USA, Austria and Japan, and four different Russian universities (located in urban and rural areas), respectively. Brandts, Saijo and Schram (2004) replicate the analysis in Japan, Netherlands, Spain and USA using the *contribution function* approach. The distribution does not significantly differ across countries, methodologies, and rural or urban areas.

#### **4-Social preferences and sanctions**

Conditional cooperators follow a solid behavioral rule: they mimic the behavior of the other group members. They adjust their decisions to the decisions of the others both upwards (when they are contributing below the average contribution of the others) and downwards (when they are contributing more than the others). Both adjustment processes could be associated with a mix of other regarding and selfish motivations, in a repeated environment. Typically, the mechanism is asymmetric: upward adjustments tend to be less intense and slower than downward changes. As Gächter (2007) puts it, 'voluntary cooperation is fragile': while conditional cooperators react to the contributions of the other team members, free riders do not. As a standard conditional cooperator is prone to apply some self-serving discount, not fully matching others' contribution, the consequence will be a decline in contributions.

The existence of a significant fraction of conditional cooperators in experimental social dilemmas raises some questions. Could cooperators get engaged in further actions to make free riding more costly? Could they conceivably find a way to anticipate the decline making free riding more costly to free riders? The experimental literature shows that ostracizing and partner selection is widely used when available, when exclusion of free riders is possible. Gächter and Thöni (2005) and Gunnthorsdottir et al. (2007) find that sorting subjects by their cooperation levels significantly enhances public good provision. de Oliveira, Croson and Eckel (2009) inform subjects about the distribution of types in their own group, being types defined on the basis of a previously run experiment. Relative to a control treatment contribution, with no information about types, contribution is significantly higher when conditional cooperators are informed about the absence of free riders. Groups with no free riders are not able to sustain cooperation. Page et al. (2005), and Cabrera et al. (2010) regroup subjects depending on their behavior. Regrouping is endogenous in the former (subjects choose whom to be matched with) while institutional in the latter (each round, the worst performer in group A is transferred to group B, while the best performer is transferred from B to A). Contribution goes up to about 70% of the social optimum in the first paper, and to almost 90% of full contribution in the second (both in group A and B!). The participants' possibility of choosing whom they are grouped with has a noticeable positive effect on efficiency.

The possibility of sanctioning free riders was first introduced in social sciences by social psychologists, as Yamagishi (1986, 1988), and political scientist, Ostrom et al. (1992); Fehr and Gächter (2000) is the canonical paper in economics. In this study, participants in a linear public good experiment get the chance of sending costly punishment points to each other after getting information about individual contributions. Punishment points are costly for the sender, and reduce receiver's earnings using a proportional rule. Rather than investing in a punishment fund to sanction the individual with the lowest contribution to the public good (as in Yamagishi, 1986), participants are allowed to send punishment points to any group member.

The difference is critical to understand how other regarding preferences shape punishment decisions. Subjects in Fehr and Gächter (2000) punish significantly more to those subjects contributing below the group average. But, some subjects do punish contributors, even when it is impossible to identify punishers (subjects are informed about the total number of punishment points they get per round). The effect of this additional stage generates a large contribution increase. More interestingly, the positive effect on public good provision is not related to possible future gains, as authors tests the same mechanism in a partners (subjects stay in the same group for the ten rounds) and strangers (groups are reshuffled every round) matching design, finding similar results.

For selfish players, punishing can hardly be a rational strategy. In the stage game of the partners condition, punishment becomes a second order public good, as subjects cannot be excluded from their potential cooperation gains. Sanctioning in the strangers condition makes even less sense for selfish players. The interesting lesson from this design is not that the possibility of excluding free riders generates cooperation gains (Croson et al., 2006 document this result in a variety of institutions and games). It is that some kind of other regarding preferences is driving a major change in decisions and group performance. As Fatas et al. (2010c) show, vertical punishment does not generate the same positive effect on contributions when it is implemented institutionally.

The effectiveness of this combination of social preferences and punishment opportunities is robust to parameter manipulations. Nikiforakis and Normann (2008) run a comparative static analysis of contribution gains as a function of the price of punishment (the receiver's earning reduction over the sender's cost). Even for low level of effectiveness, the usual decline in contributions is reversed. Bochet, Page and Putterman (2006) test the effectiveness of punishment relative to communication in a repeated social dilemma game. Punishment supplements anonymous communication and fosters contributions to almost full efficiency levels.

However, the effectiveness of sanctions is severely reduced when the organizational structure is incomplete. Fatas et al. (2010a) show that punishment is much less effective in incomplete networks, in which participants cannot

observe the full distribution of individual contributions, so punishment is sometimes not available. Moreover, it is the organizational structure, and not its density, what drives this result: star networks, in which one group member is commonly observed by all subjects, outperform any other incomplete network and does as well as a complete network (the organizational structure used in the seminal paper of Fehr and Gächter, 2000). Their results can be rationalized in terms of social preferences. In Fatas et al. (2010b) a simple model based on the existence of conditional cooperators explains how incomplete non-hierarchical networks (like the circle or the line) do worse: the signal coming from the commonly observed agent is missing.

Eckel et al (2010) document that status plays a counterintuitive role in incomplete networks (as the star). Participants mimic too closely the behavior of high status central players and don't challenge their decisions. As a consequence, punishment is rarely used, and contribution gains are negligible relative to an alternative treatment in which the status of the central player is low (her decisions are challenged openly, punishment is used, contributions increase).

The link between social preferences and punishment is still only partially understood. The existence of socially oriented subjects (e.g. conditional cooperators) does not guarantee that a socially convenient outcome will be obtained. Nikiforakis (2008) explores the effect of introducing an additional stage in which subjects may counter punish. Subjects make a contribution decision and punish; then they are able to send again punishment points, to those who punished them, or to any other participant in their group. The possibility of being punished back by free riders diminishes the use of punishment by cooperators, and its effectiveness. Hermann et al. (2008) document the wide spread of antisocial punishment in experiments run in sixteen different laboratories in the world. In almost all locations, subjects contributing above the average group contribution are punished. In some locations, the proportion of antisocial punishers is so high that the positive effect of punishment is fully lost.

## **6-Conclusions.**

Experimental economists, as well as other social scientists, have created an enormous amount of evidence documenting various dimensions of sociality in humans. In fact, mainstream economists were for a long time very skeptical about results of this type. However, the ease of replication of experiments has made it possible to investigate social behavior very carefully. Through this process many skeptics have now been convinced. It turns out that pure altruism is probably not a central ingredient of humans' pro-social motivations. Conditional cooperation and the willingness to incur costs in order to sanction non-cooperators are the two main pillars of human cooperation.

## 7-References

Anderson, S., J. Goeree and C. Holt (1998), "A theoretical analysis of altruism and decision error in public goods games", Journal of Public Economics, 70, 297-323.

Andreoni, J. (1995), "Cooperation in public-goods experiments: kindness or confusion?", American Economic Review, 85, 891-904.

Andreoni, J. (1988), "Why free ride? Strategies and learning in public goods experiments", Journal of Public Economics, 37, 291-304.

Bochet, O., Page, T., and Putterman, L. (2006), "Communication and punishment in voluntary contribution experiments", Journal of Economic Behavior and Organization, 60, 11-26.

Blount, S. (1995), "When social outcomes aren't fair: the effect of causal attributions on preferences", Organizational Behavior and Human Decision Processes, 63, 131-144.

Bolton, G.E and Ockenfels, A. (2000), "ERC: A theory of equity, reciprocity and competition", American Economic Review, 90, 166-193.

Brandts, J., Saijo, T., and Schram, A. (2004), "How universal is behavior? A four country comparison of spite, cooperation and errors in voluntary contribution mechanisms", Public Choice, 119, 381-424.

Burlando, R., and Guala, F. (2005), "Heterogeneous agents in public goods experiments", Experimental Economics, 8, 35-54.

Cabrera, S., E. Fatas, J. A. Lacomba and T. Neugebauer (2010), "Vertically splitting a firm: promotion and relegation in a team production experiment", Discussion Papers on Economic Behavior, University of Valencia.

Camerer, C. (2003), "Behavioral Game Theory: Experiments on Strategic Interaction", Princeton: Princeton University Press.

- Cameron, L. A. (1999), "Raising the stakes in the ultimatum game: Experimental evidence from Indonesia", Economic Inquiry, 27, 47-59.
- Charness, G., and Rabin M. (2002), "Understanding Social Preferences with Simple Tests", Quarterly Journal of Economics, 117, 817-869.
- Charness, G., and Yang, C. (2007), "Endogenous group formation and public goods provision: exclusion, exit, mergers and redemption", Working paper, Department of Economics, University of California-Santa Barbara. Available. at SSRN: <http://ssrn.com/abstract=932251>.
- Chaudhuri, A. (2009). *Experiments in economics: playing fair with money*. London: Routledge.
- Chaudhuri, A. (forthcoming), "Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature", Experimental Economics.
- Chaudhuri, A., and Paichayontvijit, T. (2006), "Conditional cooperation and voluntary contributions to a public good", Economics Bulletin, 3, 1-14.
- Cinyabuguma, M., Page, T., and Putterman, L. (2006), "Cooperation under the threat of expulsion in a public goods experiment", Journal of Public Economics, 89, 1421-1435.
- Cox, J., D. Friedman and S. Gjerstad (2007), "A Tractable Model of Reciprocity and Fairness" Games and Economic Behavior, 59(1), 17-45
- Cox J., D. Friedman and V. Sadiraj (2008), "Revealed Altruism", Econometrica, 76(1), 31-69.
- Crosan, R., E. Fatas and T. Neugebauer (2006): "Excludability and contribution: A laboratory study in team production", University of Luxemburg working paper.
- Dawes, R., J. McTavish and H. Shaklee (1977), "Behavior, communication and assumption about other people's behavior in common dilemma situations", Journal of Personality and Social Psychology, 35, 1-11.
- de Oleveira, A., Crosan, R., and Eckel, C. (2009), "One bad apple: uncertainty and heterogeneity in public good provision", Working paper, Department of Resource Economics, University of Massachusetts-Amherst. .
- Dufwenberg, M., and Kirchsteiger, G. (2004), "A theory of sequential reciprocity", Games and Economic Behavior, 47, 268-298.
- Eckel, C., E. Fatas and R. Wilson (2010), "Cooperation and Status in Organizations", Journal of Public Economic Theory, 12, 4, 737-762.
- Ehrhart, K.-M., and Keser, C. (1999), "Cooperation and mobility: on the run", Working paper, CIRANO and University of Karlsruhe.
- Falk, A., and Fischbacher, U. (2006), "A theory of reciprocity", Games and Economics Behavior, 54, 293-315.
- Fatas, E., M. A. Melendez-Jimenez and H. Solaz (2010a): "Punishment, Cooperation and Networks", Discussion Papers on Economic Behavior, University of Valencia.
- Fatas, E., M. A. Melendez-Jimenez, A. Morales and H. Solaz (2010b): "Public Goods and Decays in Networks", Discussion Papers on Economic Behavior, University of Valencia.
- Fatas, E., A. Morales and P. Ubeda (2010b): "Blind Justice: An Experimental Analysis of random punishment in team production", Journal of Economic Psychology, 31, 358-373.

Fehr, E., and Fischbacher, U. (2002), "Why Social Preferences Matter. The Impact of Non-selfish Motives on Competition, Cooperation, and Incentives", Economic Journal, 112, C1-C33.

Fehr, E., and Gächter, S. (2000), „Cooperation and punishment in public goods experiments”, American Economic Review, 90, 980–994.

Fehr, E., and Schmidt, K. (1999), "A theory of fairness, competition, and cooperation", Quarterly Journal of Economics, 114, 817-868.

Fehr, E., and Schmidt, K. (2006), "The economics of fairness, reciprocity and altruism - experimental evidence and new theories", in: Handbook on the Economics of Giving, Reciprocity and Altruism, Vol. 1, Serge-Christophe Kolm and Jean Mercier Ythier (eds.), Amsterdam: Elsevier, 615–91.

Fischbacher, U., Gächter, S., and Fehr, E. (2001), "Are people conditionally cooperative? Evidence from a public goods experiment", Economics Letters, 71, 397-404.

Fischbacher, U., and Gächter, S. (2009), "On the behavioral validity of the strategy method in public good experiments", Discussion Paper No. 2009-25, Centre for Decision Research and Experimental Economics, University of Nottingham.

Fischbacher, U., and Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public good experiments", American Economic Review, 100, 541–556.

Forsythe, R., Horowitz, J., Savin, N. and Sefton, M. (1994), "Fairness in simple bargaining Experiments", Games and Economic Behavior, 6, 347-369.

Gächter, S. (2007), "Conditional cooperation. Behavioral regularities from the lab and the field and their policy implications", in B. Frey and A. Stutzer (Eds.), Economics and psychology. A promising new cross-disciplinary field. CESifo Seminar Series. Cambridge: MIT Press.

Gächter, S., and Thöni, C. (2005), "Social learning and voluntary cooperation among like-minded people", Journal of the European Economic Association, 3, 303–314.

Guillem, P., E. Fatas and P. Brañas (2010), "Inducing efficient conditional cooperation patterns in public goods games. An experimental investigation", Journal of Economic Psychology, in press.

Gunnthorsdottir, A., Houser, D., and McCabe, K. (2007), "Disposition, history and contributions in public goods experiments", Journal of Economic Behavior and Organization, 62, 304–315.

Güth, W., Schmittberger, R., and Schwarze, B. (1982), "An Experimental Analysis of Ultimatum Bargaining," Journal of Economic Behavior and Organization, 3, 367-388.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Gintis, H., McElreath, R., and Fehr, E. (2001), "In search of homo economicus: experiments in 15 small-scale societies", American Economic Review, 91, 73-79.

Hermann, B., and Thöni, C. (2008), "Measuring conditional cooperation: a replication study in Russia", Experimental Economics, 12, 87–92.

Hermann, B., Thöni, C., and Gächter, S. (2008), "Antisocial punishments across societies", Science, 319, 1362–1367.

Isaac, M., and Walker, J. (1988), "Communication and free riding behavior: the voluntary contributions mechanism", Economic Inquiry, 26, 585-608.

Kagel, J., and Roth, A. (1995), Handbook of experimental economics, Princeton: Princeton University Press.

- Keser, C., and van Winden, F. (2000), "Conditional cooperation and voluntary contributions to public goods", Scandinavian Journal of Economics, 102, 23-39.
- Kocher, M., Cherry, T., Kroll, S., Netzer, R., and Sutter, M. (2008), "Conditional cooperation on three continents", Economics Letters, 101, 175-178.
- Kurzban, R., and Houser, D. (2005), "An experimental investigation of cooperative types in human groups: a complement to evolutionary theory and simulations", Proceedings of the National Academy of Sciences, 102, 1803-1807.
- Ledyard, J. (1995), "Public goods: some experimental results", in J. Kagel and A. Roth (Eds.), Handbook of experimental economics, Princeton: Princeton University Press (Chap. 2).
- Nikiforakis, N. (2008), "Punishment and counter-punishment in public good games: can we really govern ourselves?", Journal of Public Economics, 92, 91-112.
- Nikiforakis, N., and Normann, H. (2008), "A comparative statics analysis of punishment in public good experiments", Experimental Economics, 11, 358-369.
- Offerman, T., Sonnemans, J., and Schram, A. (1996), "Value orientations, expectations and voluntary contributions in public goods", Economic Journal, 106, 817-845.
- Ostrom, E., Walker, J., and Gardner, R. (1992), "Covenants with and without a sword: self-governance is possible", American Political Science Review, 86, 404-417.
- Page, T., Putterman, L., and Unel, B. (2005), "Voluntary association in public goods experiments: reciprocity, mimicry, and efficiency", Economic Journal, 115, 1032-1053.
- Palfrey, T. and Prisbrey, J. (1997), "Anomalous behavior in linear public goods experiments: how much and why?", The American Economic Review, 87, 829-846.
- Rabin, M. (1993), "Incorporating fairness into economics and game theory", American Economic Review, 83, 1281-1302.
- Rotemberg, J. (2006), "Altruism, Reciprocity and Cooperation in the Workplace", in Serge-Christophe Kolm and Jean Mercier Ythier, (eds.), Handbook on the Economics of Giving, Reciprocity and Altruism vol. 2, North Holland, 2006, 1371-1407.
- Sonnemans, J., Schram, A., and Offerman, T. (1999), "Strategic behavior in public good games: when partners drift apart", Economics Letters, 62, 35-41.
- Slonim, R., and Roth, A. (1998), "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic", Econometrica, 66, 569-596.
- Vesterlund, L. (2010), "Public goods experiments", to appear in The Handbook of Experimental Economics, vol 2. (J. Kagel and A. Roth, eds.).
- Yamagishi, T. (1986), "The provision of a sanctioning system as a public good", Journal of Personality and Social Psychology, 51, 110-116.
- Yamagishi, T. (1988), "The provision of a sanctioning system in the United States and Japan", Social Psychology Quarterly, 51, 265-271.