

## **Micromotives, Microstructure, and Macrobehavior: The Case of Voluntary Cooperation**

**Simon Gächter**

School of Economics, University of Nottingham, Nottingham, UK

**Christian Thöni**

Department of Economics, University of St. Gallen, St. Gallen,  
Switzerland

*How micromotives, the microstructural features of interactions, and macrobehavior are related is a fundamental question in all social sciences. In this article we argue that laboratory experiments are a useful tool to study this question, because the experimenter can measure motivations, manipulate microstructures, and sometimes even exploit variation in the macrosocial environment. We illustrate the experimental approach with the help of four examples from the context of voluntary cooperation. The examples (from previous and new experiments) illustrate the link between individual pro-social motives and wage formation in labor markets, the inevitability of the “Tragedy of the Commons,” and two potential ways to avoid the tragedy: sorting and punishment.*

**Keywords:** gift exchange, laboratory experiments, macrobehavior, macrosocial background, micromotives, microstructure, punishment, reciprocity, sorting, voluntary cooperation

### **1. INTRODUCTION**

How individual decisions influence aggregate-level outcomes is a fundamental question in the social sciences. In modern societies, millions of individuals take decisions about what to buy or sell, and their decisions influence the price at which trading will occur.

We gratefully acknowledge financial support by the University of Nottingham and the Grundlagenforschungsfonds of the University of St. Gallen. We also benefited strongly from the excellent comments by the anonymous referees, and by Vincent Buskens, Werner Raub, and Marcel van Assen.

Address correspondence to Simon Gächter, Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham, Sir Clive Granger Building, University Park, Nottingham, NG7 2RD, United Kingdom. E-mail: simon.gaechter@nottingham.ac.uk

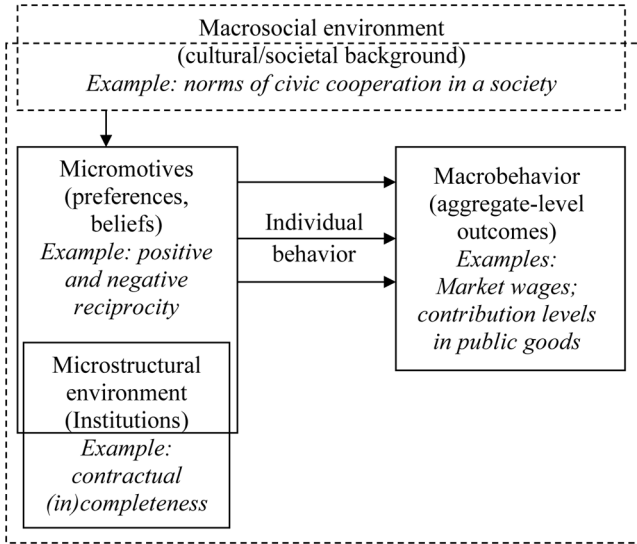
Similarly, people take decisions which, collectively, have consequences for the (global) environment. The extent of people's civic engagement determines the quality of neighborhoods, communities, and political entities. Social interactions will also influence the success of firms and other organizations.

There are a variety of tools social scientists have developed to understand the links between people's preferences and beliefs (micromotives), the structure in which people make decisions and with whom they interact (microstructure, often called institutions or rules of the game), and the aggregate-level outcomes (macrobehavior). Rational choice theory, and in particular *game theory*, are probably the most important theoretical tools because they allow us to model assumptions about micromotives and microstructures to predict outcomes (Schelling, 1978; Coleman, 1990; Hechter and Kanazawa, 1997; Gintis, 2009; Buskens and Raub, 2008).

A further tool to investigate the link among motives, structures, and outcomes is *simulations* and *agent-based modeling*, in particular if the model is too complex to solve analytically (Macy and Willer, 2002). These tools have been used for explaining several phenomena of interest in this article (e.g., Boyd, Gintis, Bowles, and Richerson, 2003; Boyd, Gintis, and Bowles, 2010; Helbing and Yu, 2009).

In this article we argue that in addition to theory and simulations *laboratory experiments* are a suitable complementary empirical tool. In experiments we can observe real rather than simulated behavior, and we do so in a controlled decision environment that allows *causal* inferences about the variables of interest (Smith, 1982).<sup>1</sup> That is, the experimenter can design experiments that shed light on people's motivations and can then see how these micromotives shape an aggregate-level outcome of a social unit of interest (the macrobehavior), given a particular decision situation (a microstructure, such as a market environment, or one-shot or repeated interactions). Similarly, the experimenter can change the microstructure and thereby get insights into the causal consequences of particular microstructures, holding people's micromotives constant. Using cross-cultural experiments, researchers can also study the impact of the macrosocial (societal) background on people's behavior and its consequences for aggregate-level outcomes of interest. Figure 1, which is inspired by Coleman (1990, ch. 1) and adapted to our purposes, illustrates our

<sup>1</sup>See Carpenter, Harrison, and List (2005); Levitt and List (2008); Falk and Heckman (2009); Bardsley et al. (2010); and Rosser and Eckel (2010) for recent discussions of the methodology of experimental economics. The edited volume by Webster and Sell (2007) contains several articles on methodology and practicalities of experiments in sociology.



**FIGURE 1** The relationship between micromotives, microstructures, and the resulting macrobehavior in a given macrosocial environment.

approach schematically. Micromotives and the details of the microstructural environment jointly determine individual behavior. The macrosocial environment might also influence people's micromotives. Individual behavior shapes the macrosocial outcomes.

While we will use Figure 1 as a framework (rather than as a theory) to illustrate how experiments can shed light on micro-macro relationships, we should add that the scheme can equally be used to discuss theoretical and simulation approaches to the same issue. For example, rational choice approaches make assumptions about people's preferences and beliefs, as well as the microstructure in which people operate and then derive predictions for the macrobehavioral outcome, typically by using game-theoretic tools. If analytical conclusions are infeasible, simulations may prove of great help. This description should not create the impression, however, that these approaches are three disconnected tools. In particular, experiments and theory are closely connected. Most (economic) experiments (and all examples discussed here) use game theory to design the experiment and to derive theoretical benchmark predictions against which empirical results are compared. Moreover, testing theories and providing input into new theories is a very important task of economic experiments (for recent methodological discussions on this issue see references in

footnote 1). Simulations and experiments are less closely related than theory and experiments, at least in experimental economics. However, behavioral regularities uncovered in experiments can be used as input into simulations and sometimes simulations are used to derive hypotheses of what might happen in an experiment (see, e.g., Carpenter, 2007).

### 1.1. One Theme—Four Examples

We will illustrate our approach with one theme, and four different examples. The theme is “voluntary cooperation,” by which we mean people’s readiness to act against their immediate self-interest with beneficial consequences for others (and maybe oneself in the future). The examples discuss wage formation, the inevitability of the “Tragedy of the Commons” (Hardin, 1968), and two potential ways out of it: sorting and punishment. Three examples will investigate the roles of different micromotives and microstructures for macrobehavior, holding the macrosocial environment (the cultural or societal background) constant. The fourth example will utilize variation in the macrosocial environment in which the experiments have been conducted. As we will see, the common element of all four examples is heterogeneity in people’s micromotives. This heterogeneity has different consequences for macrobehavior, depending on the microstructural environment, and, in our fourth example, also depending on the details of the macrosocial environment.

Voluntary cooperation and the micromotives behind it are interesting because voluntary cooperation is often necessary to avoid potential inefficiencies of contractually incompletely specified situations. A situation is contractually incompletely specified if the collectively efficient outcome cannot be achieved by means of an enforceable contract. Preventing climate change is a striking example, but the problem occurs at much smaller scales as well. Arguably, contractual incompleteness is a generic feature of a great many social interactions (Bowles, 2003) and is therefore also intimately linked to the problem of social order.<sup>2</sup> Socially beneficial agreements need to be self-enforcing, and the question arises how individual motives and institutions facilitate self-enforcement.

<sup>2</sup>See Hechter and Horne (2009) for a collection of important texts on the problem of social order. The problem of social order is also discussed in Fehr and Gintis (2007) and Buskens and Raub (2008). Buskens and Raub point out, for example, that already Durkheim and Weber saw contractual incompleteness as a central issue.

For a long time social theorists, in particular in economics and biology, tended to explain cooperation (or the conditions under which cooperation can flourish) by characteristics of the microstructural environment that give self-regarding individuals an incentive to cooperate (see Sigmund, 2010, for a recent survey of the relevant models). However, while the focus on selfishness and microstructures has produced invaluable insights into the problems of voluntary cooperation, research in the last two decades has considerably challenged the view that individuals are largely self-regarding. Many people are willing to incur costs to help those who helped them, or to punish wrongdoers even in one-shot situations. This motivation has been called “strong reciprocity” (Gintis, 2000). This is not to say that people are not self-interested at all. In most experiments, a considerable fraction of people are best described as self-interested, but a majority of people display strong reciprocity. Strong reciprocity and material self-interest are the most important micromotives that underlie the examples we discuss in this article.

The first two examples illustrate how the existence of strong positive reciprocity can lead to very different macrobehavioral outcomes, depending on the microstructural environment that influences the details of people’s interactions. Specifically, the first example is taken from labor markets and illustrates the role of voluntary cooperation on wage formation. The micro-level motives we look at are “gift exchange,” which is an example of strong positive reciprocity whereby people reciprocate fair wages with high effort, and profit maximization. We will show that both motives have a very strong influence on aggregate market outcomes (wages). This outcome is radically different if the microstructural interaction is governed by complete contracts (i.e., the efficient contract is implemented by design). In this case, wages converge quickly to an outcome where gift exchange seems entirely unimportant. Thus, this experiment illustrates how micromotives influence macro-level outcomes in one particular microstructure (institution) and how the macro-level outcome changes if the institutional rules change.

The second example is the simplest one in the sense that we keep the microstructural setup constant. The experiment investigates the relationship between people’s individual pro-social attitudes (which can also be seen as a case of strong positive reciprocity) and the resulting aggregate cooperation level, measured as contributions to a public good. This experiment illustrates that under conditions in which group members are randomly matched and punishment is not possible, the Tragedy of the Commons, that is, the breakdown of cooperation, is almost inevitable. This is true despite the fact that

most people are not free rider types but reciprocators (“conditional cooperators”). The comparison to the first example shows that contractual completeness is not necessary but only sufficient for a macrobehavioral outcome that resembles the one predicted by selfishness; such a macrobehavioral outcome is also possible if contracts are incomplete.

The first two examples illustrate the joint importance of positive reciprocity and the details of interaction structures for voluntary cooperation. This observation raises the question what would happen if people would not interact in randomly composed groups but in groups of like-minded cooperators. Specifically, in the third example we address the question whether people who share similar attitudes of strong positive reciprocity behave differently when they know that they are among like-minded others. For instance, can cooperators who know that others are cooperators as well avoid the Tragedy of the Commons? For answering this question, we will present the results of new experiments. In these experiments we will measure cooperative attitudes (micromotives) and vary some relevant microstructural conditions. We will show that the answer to our question is yes, provided it is supported by microstructural conditions that make imitation of cooperators by free riders difficult.

The first three examples illustrate interaction effects of microstructural environments with many people’s micromotive of *positive* reciprocity. The final example turns attention to *negative* reciprocity (a.k.a. punishment) and its potential to sustain cooperation. The main purpose of this example is not to show that negative reciprocity can support cooperation (this has been shown before; in the context of discussing the problem of social order see, e.g., Fehr and Gintis, 2007). Rather, it is to show that the macrosocial environment can exert a very strong influence on people’s micromotives of punishment which, in turn, lead to very different aggregate cooperation levels, despite the fact that all experiments were conducted in exactly the same microstructural environment.

## 1.2. Methodological Considerations

In addition to providing four substantive examples on behavioral insights into the micro-macro link, our article also has a methodological message. We believe that *laboratory* experiments are a particularly apt tool to study the links depicted in Figure 1 not only theoretically but also behaviorally. The main reason for this assessment is that the laboratory offers a degree of control that other empirical methods, including field experiments, can hardly achieve.

This is particularly true with regard to the detailed measurement of individual preferences and beliefs and the testing of the *causal* impacts of micromotives and microstructures, and the interaction of micromotives and microstructures. Our emphasis on lab experiments does not mean there is no role for field experiments, that is, experiments that are conducted in more naturally occurring environments. However, natural field experiments, in which participants do not even know that they participate in an experiment (Carpenter, Harrison, and List, 2005), mostly do not permit the detailed measurement of individual motives and beliefs, as well as the comprehensive *ceteris paribus* variation in the microstructural environment. The reason is that these measurements and variations require more experimental intervention than is feasible within a natural field experimental situation where people by definition are not aware that they participate in an experiment.

There are a couple of further concerns with lab experiments which we would like to address before we discuss our substantive examples in the main part of the article. The first concern is that experiments with undergraduates lack representativeness, and the second concern is whether the results obtained under the artificial conditions of the lab can be generalized to the “real world.” It is, of course, correct to observe that students are not representative. However, whether this is a problem depends on the exact research question. If the research question is to test a theoretical claim, students are actually an excellent subject pool: (i) decision theories typically do not come with caveats that the theories do not hold for students, and (ii) because students are educated they likely understand the rules of the experiment and therefore the results are not confounded with lack of understanding of the decision situation. To extend this argument to our current purposes, we see our examples as *behavioral models* of specific micro-macro links; therefore, we should be able to gain *qualitative* insights into this relationship with any subject pool, and hence also with students. Moreover, we expect that the most important differences between subject pools lies in the distribution of people’s micromotives—older cohorts are typically less selfish than younger ones (we discuss some of this evidence as we go along). This is a quantitative issue and the qualitative insights should not be affected much (this is, of course, a testable claim).

A second frequent issue that skeptics raise against experiments concerns their apparent lack of realism; that is, the objection that experiments are conducted in artificial lab environments that do not occur in real life and therefore generalizing to reality is difficult.

We believe this argument is often misguided. First, experiments indeed lack realism if an experiment would be taken as a literal description of a real situation. Yet, like theoretical models, experiments are abstractions, that is, *models* of generic features of naturally occurring situations. This is true of the gift exchange game introduced in the next section, which models the generic features of contractual incompleteness in employment relationships. It is also true of the public goods game as a model of multilateral cooperation (discussed in examples 2–4). Second, in economic experiments real people take decisions with real consequences (because earnings in the experiment depend on decisions) and in this sense the laboratory situation is as real as any other naturally occurring situation. Third, most naturally occurring situations typically take place at a particular time and in a particular location and are therefore often no less special in their specific configuration than lab situations. Lab experiments have the added advantage that (an almost) exact replication of a decision situation is feasible (the latter is less true in field experiments, where replication is often infeasible).

As Falk and Heckman (2009) point out, behavior in both the laboratory and any naturally occurring situation is a function of all contextual variables that define the respective situation. Hence, the real issue is how best to isolate the causal effect of interest. For example, testing the “broken windows theory” of norm violations might best be done in a field experiment (see Keizer, Lindenberg, and Steg, 2008) because the necessary control conditions can hardly be implemented in a lab environment. The causal effects we discuss in this article are best isolated in a lab experiment, because they require detailed measurement of micromotives and full control of microstructures hardly available in naturally occurring contexts (see Croson and Gächter, 2010, for a more general discussion of this tradeoff). In general, isolating causal consequences of a change in one variable requires holding all other potential influence factors constant. Thus, in principle, any claim about causal effects always depends on the factors that are held constant. This is true of lab as well as field experiments. Therefore, if a field situation is compared to a structurally similar lab situation (see, e.g., Benz and Meier, 2008), the test is whether the additional factors inevitably present in the richer field setting matter for behavior. If a phenomenon does not occur in a field experiment but in a lab experiment, this cannot be taken as evidence against the lab results (provided the lab findings are replicable). In our view the only scientifically productive question is which additional factors in contextually inevitably richer field settings are responsible for the different result.



### 1.3. Our Contribution

This article is related to previous surveys on experimental findings that speak to sociological questions (Fehr and Gintis, 2007; Buskens and Raub, 2008). Fehr and Gintis discuss the problem of social order and issues surrounding social exchange in more depth than we do. Buskens and Raub (2008) concentrate mainly on the bilateral trust game and how “embeddedness” in repeated dyadic relationships and in networks changes behavior. While our focus is on laboratory experimental methods, Buskens and Raub also discuss other nonexperimental research methods. Furthermore, in contrast to both papers we not only survey the existing literature but also provide evidence (presented in Section 4) from new experiments on assortative matching, which is not discussed by the other papers. Finally, we provide an update on some of the more recent literature.

## 2. GIFT EXCHANGE, PROFIT MAXIMIZATION, AND ITS MACROCONSEQUENCES

Our first example on the micro-macro link is taken from labor economics. In a seminal paper, Akerlof (1982) argues that the employment relation is best viewed as a “gift exchange”; that is, employers and employees voluntarily exchange work effort for good working conditions and payment. The reason for this view is that the employment relation is an incomplete contract, which leaves many important aspects, in particular work effort, unregulated and therefore nonenforceable by third parties (Bowles, 2003). However, contractual incompleteness gives agents an incentive to shirk, which may lead to an inefficiently low surplus.

How to initiate voluntary cooperation? Akerlof (1982) argued that firms might have an incentive to offer their employees good working conditions or above-market wages to induce gift exchange. Norms of reciprocity (e.g., Gouldner, 1960) might induce workers to supply effort above levels self-interested employees would choose. Suppose contractual incompleteness is prevalent and many firms have an incentive to pay wages above market clearing levels. Put differently, profit-maximizing firms might not be willing to lower wages because they fear that this would undermine work morale. The aggregate (macro-level) consequence of the micromotives of gift exchange and profit maximization is that wages might be rigid and above market-clearing levels. If that happens, involuntary unemployment will be an inevitable consequence.

Akerlof (1982) developed his theoretical argument on the basis of sociological case studies and social psychological insights. However,

based on such data it is difficult to prove whether gift exchange can indeed *cause* wage rigidity. A couple of difficulties need to be overcome. First, in naturally occurring labor relations, what *looks* like gift exchange might actually be motivated by strategic or reputational motives: I work hard because otherwise I expect adverse consequences (lower wages, worse conditions, not being promoted, or even being fired). Would nonselfish, genuinely reciprocal gift exchange also work in the absence of strategic motives? Second, and related, in naturally occurring situations it is difficult (if not impossible) to tell what exactly self-interested behavior predicts and what are market clearing wages. Without this knowledge it is hard to say whether there is wage rigidity and whether it is *caused* by gift exchange and profit maximization. A laboratory experiment is necessary. In the following, we describe the experiments by Fehr, Kirchler, Weichbold, and Gächter (1998) because this study investigates the impact of contractual incompleteness and gift exchange on wage formation.

## 2.1. Experiments and Results

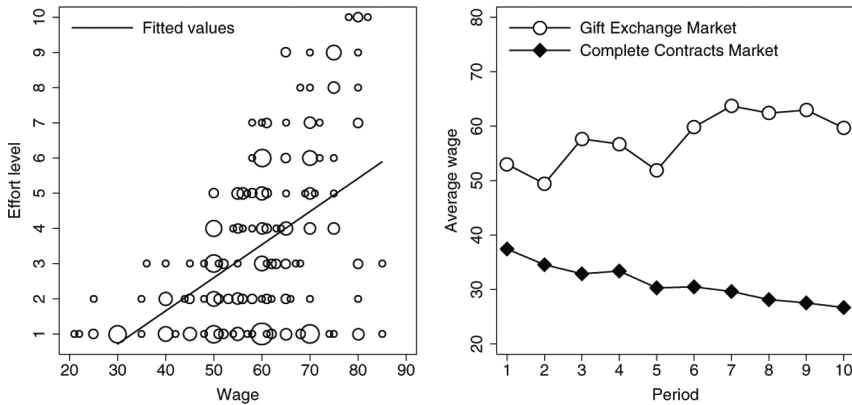
The work horse for studying the micro-macro link is a version of the gift exchange game developed by Fehr, Kirchsteiger, and Riedl (1993). In the gift exchange game there are two roles, employers and employees. Each employer can only hire one employee, and there are more employees than employers. The sequence of events is as follows. Subjects in the role of an employer make wage offers in a competitive market institution known as “one-sided auction” (see Davis and Holt, 1993, for a detailed description). Wages are between 20 and 120 monetary units. Employees see these wage offers and can accept any wage offer that is still available. Acceptance of a wage offer concludes an employment contract. Notice that contracts are incomplete because employers can only offer a fixed wage but cannot specify a particular effort, so effort is not contractible. In the next step employees choose their effort and the game ends. There are 10 different effort levels. “Effort” in this experiment means choosing a number with the consequence that the higher the chosen number the higher is the employer’s profit and the higher are the employee’s effort cost. The payoff of employers increases in the effort of the employee and decreases in the wage paid to the employee. The employees’ payoffs increase in wage and decrease in effort. Parameters are such that maximal effort would maximize the total surplus available. Thus, the gift exchange game can be seen as a generalized version of a sequential Prisoner’s Dilemma or trust game (discussed in Buskens and Raub, 2008, and in Vieth, 2009).

The experiment is conducted under anonymity and the market interaction described above is repeated for 10 periods. In the framework of Figure 1, the complete set of rules, which are explained in detail to the subjects, constitutes the microstructural environment in this experiment. This setup ensures that there are no strategic reasons for gift exchange. If for people's micromotives we assume that all players are rational and self-interested payoff maximizers, then employees will, irrespective of the wage, choose the minimum effort because effort is costly. Employers therefore have an incentive to pay the lowest wage because higher than minimal wages cannot trigger gift exchange from selfish employees. Since there are fewer employers than employees, employers are in a stronger position and should be able to push wages down to the lowest possible level of 20. Thus, low wages would constitute the macrobehavioral outcome. However, if sufficiently many people are motivated by strong positive reciprocity, profit maximizing employers have incentives to offer nonminimal wages and the macrobehavioral outcome would be characterized by high wages.<sup>3</sup>

Numerous gift exchange experiments have been conducted in this framework (Fehr, Goette, and Zehnder, 2009). Figure 2 shows the results of the gift exchange experiments of Fehr et al. (1998). The left panel shows a bubble plot of the wage-effort combinations and a regression line. Overall, the self-interest prediction is clearly refuted because there is a highly significantly positive relation between wage and effort. Despite the monetary incentive to choose minimal effort, experimental subjects in the role of the employee tend to reward generous wage offers by high efforts. This is clear evidence for strong positive reciprocity (called gift exchange in this context) as an important micromotive. However, the figure also shows that there is substantial heterogeneity. Irrespective of the wage paid by the firm, there is always a fraction of minimal effort choices. *Homo economicus* and *homo reciprocans* co-exist, although the latter seem to outnumber the former (see also Gächter and Thöni, 2010).

Does gift exchange lead to nonminimal wages and wage rigidity? The right-hand panel of Figure 2 gives the answer. In the gift exchange market described above wages are indeed far above the predicted level of 20 and even increase over time. This is unambiguous

<sup>3</sup>In the experiment all players are only informed about the rules of the experiment and how decisions translate into payments. Hence, employers actually only learn about their employees' degree of reciprocity as the experiment progresses. At the beginning of the experiment employees can only form beliefs (drawn from their life experience) about how likely reciprocal reactions are.



**FIGURE 2** Left panel: Individual-level motivation (positive reciprocity). Right panel: macro-level market wages when reciprocity is possible (gift exchange market) and when it is not possible (complete contracts market). The predicted (market clearing) wage is 20. *Data source:* Fehr et al. (1998).

evidence for nonmarket clearing wage rigidity. Further analyses showed that the average wage observed in these markets was indeed the profit-maximizing wage, given the employees' average wage-effort relationship. But are gift exchange and contractual incompleteness the *cause* behind this finding? To investigate this crucial question, Fehr et al. (1998) included a further treatment in their design, called the "Complete Contracts Market." In this treatment effort is fully contractible (set exogenously at the highest effort level). Thus, the contract is complete because there is no effort discretion any longer (and there are also no possibilities for gift exchange). The results, also depicted in the right-hand panel of Figure 2, show that wages are dramatically different when contracts are complete: consistent with the selfishness prediction developed above, wages are substantially lower from the start and converge toward 20 by the end of the experiment. This confirms findings from numerous experiments conducted with complete contracts: competitive markets converge to the market clearing equilibrium (Davis and Holt, 1993), even if the resulting distribution of the gains from trade is unequal. Thus, contractual incompleteness and strong positive reciprocity cause wage rigidity as a macrobehavioral outcome in this experiment.

One objection against these experiments might be that in reality employment relationships are not one-shot market transactions but rather long-term relations. Experiments that allow for repeated interactions as a more realistic microstructural assumption have

shown that under such conditions gift exchange is even stronger than in one-shot games and wage rigidity is reinforced (e.g., Falk, Gächter, and Kovacs, 1999).

## 2.2. Discussion

There are several potential psychological sources behind the gift exchange we observe in these experiments. Micromotivations comprise compliance with the social norm of reciprocity (Gouldner, 1960; Diekmann, 2004), equity concerns (Adams, 1965), inequity aversion (Loewenstein, Thompson, and Bazerman, 1989; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), altruism (Cox, Friedman, and Sadiraj, 2008), reciprocity (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006), guilt aversion (Charness and Dufwenberg, 2006), and social esteem (Ellingsen and Johannesson, 2008). The fact that several psychological motives can explain gift exchange suggests that, at least on average, gift exchange (or positive reciprocity in general) is an important behavioral motivation.

Given the sequential structure of interactions in the gift exchange game (contracts are concluded before effort decisions are made) profit-maximizing employers are well advised not to pay too low wages, for in this case effort levels will be low. As mentioned above, this point was already raised by Akerlof (1982) in a theoretical conjecture based on case study evidence. The experimental results provide support for Akerlof's conjecture. Notice that employers do not have to be motivated by fairness motives or any other-regarding motive to come to this conclusion; profit maximization suffices. In terms of Figure 1, as a consequence of these micro-level motivations, the macrolevel outcome will be wage rigidity at nonmarket clearing levels. If contracts are complete, wages do converge toward market clearing levels. This observation, in combination with the findings under incomplete contracts suggests two things. The microstructure of interaction (complete vs. incomplete contracts in this case) is decisive for the aggregate outcome. A second insight of these and numerous other experiments is that under complete contracts social preferences do not matter for shaping the aggregate outcomes in competitive markets. That is, although on average people are not selfish but motivated by strong positive reciprocity, markets with complete contracts are likely to converge to predictions derived under selfishness. These points have also been made clear in formalized theories of social preferences that were developed to explain, for instance, results from gift exchange experiments (see, e.g., Fehr and Schmidt, 1999). However, as our next example will show, it can also happen

that under incomplete contracts macro-level outcomes converge to predictions derived under selfishness.

### 3. THE TRAGEDY OF THE COMMONS IS (ALMOST) INEVITABLE

Another important class of situations with contractual incompleteness concerns cooperation problems, the most famous of which is the Prisoner's Dilemma. A well-known feature of this game is that mutual cooperation is the collectively best outcome, but unilateral defection is always profitable. As a consequence, rational and selfish players will end up playing mutual defection. This game models contractual incompleteness because the socially efficient outcome is not contractible. A large number of experiments have been conducted (Rapoport and Chammah, 1965; Sally, 1995; Kollock, 1998; Colman, 1999), and the Prisoner's Dilemma has been central in understanding the evolution of cooperation (e.g., Sigmund, 2010). For instance, the possibility to defect can be used as a threat to uphold cooperation, if the interaction is long enough. However, there is an important limitation to this possibility of punishment. Many human cooperation problems occur in larger groups, and in this case bilateral punishment strategies do not apply easily or not at all. For these reasons, research has turned to social dilemma and public goods games, which are effectively  $n$ -person generalizations of the Prisoner's Dilemma game. In the following we will focus on public goods games.

In a public goods game  $n$  players are endowed with a certain amount of money,  $e$ . Players simultaneously choose a contribution to the public good,  $g_i \in [0, e]$ . Player  $i$ 's payoff depends on his or her contribution as well as the contributions of the other players according to the following payoff function:

$$\pi_i = e - g_i + \frac{\alpha}{n} \sum_{j=1}^n g_j.$$

For  $\alpha > 1$  the joint-income maximizing solution requires all players to contribute their whole endowment. However, as long as  $\alpha/n < 1$ , every player has a strictly dominant strategy to contribute nothing to the public good because the individual return from the public good is less than the foregone payoff from the private good. This setup is easy to understand and therefore ideally suited for experiments. While stylized and simplified, the public goods game is a model of a great many social situations—from teamwork and neighborhood watch to environmental protection, voluntary provision of public goods, paying

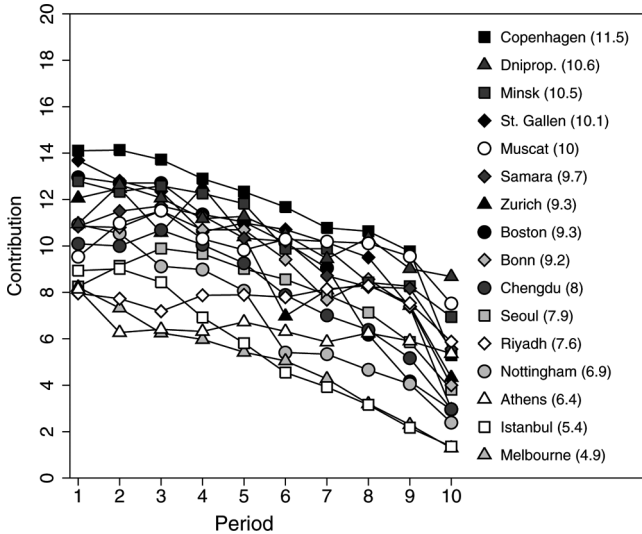
taxes, fighting corruption, common pool resource management, and so on. These situations may differ in details, but they all have in common that rationality and self-interest predict a socially inefficient outcome, that is, the Tragedy of the Commons, as Hardin (1968) called this bleak macrobehavioral outcome. However, numerous Prisoner's Dilemma and other experiments, including the gift exchange experiments discussed in the previous section, have shown that many people are not exclusively self-interested but are also motivated by strong positive reciprocity. Furthermore, as Ostrom (1990, 2005) shows, in many naturally occurring cooperation problems people manage to achieve much more efficient solutions than Hardin's view predicts. This suggests that conditions exist under which (nonselfishly motivated) people might be able to prevent the Tragedy of the Commons. Experiments are a useful tool to investigate these conditions because self-interest and collective interest, as well as the microstructure of interactions can be controlled by the experimenter.

### 3.1. Experiments and Results

In terms of Figure 1, the microstructural environment in this experiment is the public goods game described above, including further procedural details, such as parameters and interaction structures. A frequently used parameter set in experimental investigations is  $\alpha = 1.6$ ,  $n = 4$ , and  $e = 20$  monetary units. Most experiments are conducted under anonymity, and the game is repeated to study the development of cooperation. In repeatedly played public goods games participants are typically informed about what others have contributed (they learn either the average or all individual contributions). The observed cooperation levels constitute the macrobehavioral outcomes.

Figure 3 illustrates the findings of Herrmann, Thöni, and Gächter (2008a), who ran public goods experiment in 16 subject pools worldwide using the same experimental protocol and procedures everywhere to maximize comparability across subject pools.<sup>4</sup> In their experiments, stable groups of four played the public goods game for ten periods, and this was known to the participants. Thus, since the experiments were conducted in different societies, the experiment also allows gauging whether the macrosocial environment influences macrobehavior.

<sup>4</sup>The experimental instructions and all further details can be found in Herrmann, Thöni, and Gächter (2008b).



**FIGURE 3** The “Tragedy of the Commons” in the public goods game. *Data source: Herrmann et al. (2008a).*

The results obtained by Herrmann et al. (2008a) are consistent with many similar experiments reported in the literature. Initially people contribute roughly half their endowment to the public good but then cooperation unravels and in many cases reaches very low levels.<sup>5</sup> Hardin’s prediction seems to prove right. The macrosocial environment seems to have some influence on cooperation levels, but the decline of cooperation happened everywhere.

Why is the Tragedy of the Commons apparently unavoidable in this context? Is it because people’s other-regarding motives are too weak and people are actually mostly selfish and initial cooperation just occurred due to confusion? Or is it due to preference heterogeneity in the sense that some people might be cooperators, whereas others are selfish, as suggested by the results from the gift exchange game (see Figure 2, right-hand panel)? More generally, can we explain the macro-level outcome of a breakdown of cooperation by micro-level motivations? The experiment by Herrmann et al. (2008a) and others do not allow inferences about the possible micromotives behind the macrobehavioral outcome of the breakdown of cooperation.

<sup>5</sup>See Fischbacher and Gächter (2010) and the references therein.

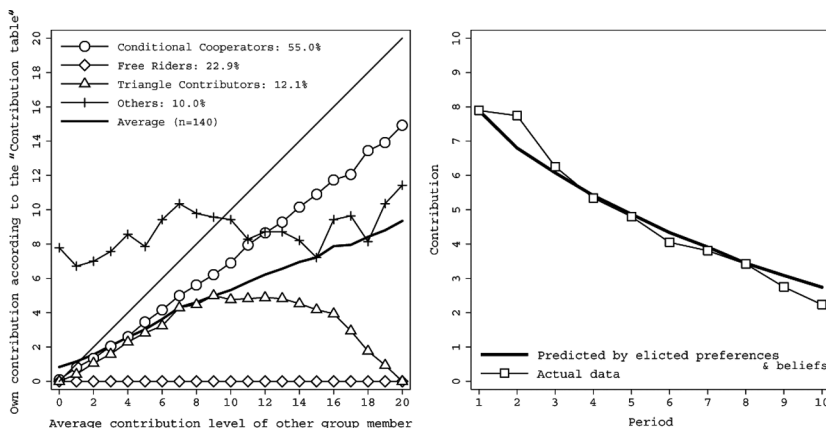


Measuring people's micromotives in a public goods game is not an entirely straightforward task. Here is why. The public goods game is a simultaneous move game. Observing a single contribution decision in an experiment provides only little information about a subject's preferences, because we usually do not know what contributions a subject expects from other players. For a money-maximizing player, beliefs about others' contributions are irrelevant because she chooses to contribute zero anyway. However, the converse argument does not hold. Observing a subject choosing a contribution of zero does not necessarily mean that this subject is maximizing her monetary income. It could be the case that this subject contributes zero because she expects others to do alike (Kelley and Stahelski, 1970; Croson, 2007). Thus, beliefs about others' contributions constitute an important element of people's micromotives for cooperation, but these beliefs might be endogenous or correlated with one's own cooperative attitude.

Fischbacher, Gächter, and Fehr (2001) introduced an experimental method to measure micromotives behind contribution decisions in public goods games in a way that avoids the endogeneity problem. In these experiments people are asked in an incentive compatible way to indicate how much they will contribute given all possible average contributions of other group members, that is, subjects are asked to fill in a "contribution table." Fischbacher et al.'s design ensures that people motivated by own payoff maximization alone have an incentive to enter 0 for all average contribution of others. "Conditional cooperators" might be willing to increase their contribution the more others contribute. Thus, micromotives are measured directly and comprehensively.

The left-hand panel of Figure 4 shows the results of a recent study (with  $n = 140$  participants) by Fischbacher and Gächter (2010), who use the same experimental design as Fischbacher et al. (2001) and obtained similar results. On the horizontal axis we depict the average contribution level of the other group members. In the public goods game used in this study subjects played the game in groups of four and received an endowment of 20 units. Therefore, the average refers to the contribution of three other subjects and takes a value between 0 and 20. On the vertical axis we depict the conditional contribution according to the contribution table.

The patterns observed by Fischbacher and Gächter (2010) suggest the existence of distinct types: (i) 22.9% of the subjects behave as money-maximizing players and contribute zero in all cases ("Free Riders"); (ii) the majority of the subjects (55%) contribute more whenever other subjects contribute more ("Conditional Cooperators"); (iii) a minority of subjects (12.1%) are "Triangle Contributors." These



**FIGURE 4** Left panel: Individual cooperation preferences. Right panel: Aggregate-level consequences of individual heterogeneity. *Data source:* Fischbacher and Gächter (2010).

subjects increase their contribution up to some point and decrease it thereafter. The remaining 10% of subjects do not fit in any of these categories.<sup>6</sup>

Can this distribution of types explain the decline of cooperation? To investigate this question Fischbacher and Gächter (2010) elicit subjects' contribution preferences in one experiment and then observe the same people over 10 periods of a standard public goods game, described above. In addition to making contributions, subjects are also asked in each period to state their beliefs about how much others will contribute. Fischbacher and Gächter then use the beliefs and the individual contribution preferences as elicited in the contribution table to predict contributions. Figure 4, right-hand panel, shows that on average actual contributions are predicted remarkably well. Econometric estimations and in particular simulations reveal that the reason

<sup>6</sup>Replication studies investigated the distribution of types in different subject pools. Herrmann and Thöni (2009) find an identical fraction of Conditional Cooperators but a smaller fraction of Free Riders in experiments conducted in four cities in Russia. Kocher, Cherry, Kroll, Netzer, and Sutter (2008) conducted the experiment in Austria, Japan, and the United States and found a substantially higher share of conditional cooperators in the United States than in the other countries. Finally, Thöni, Tyran, and Wengström (2009) report results from a large-scale Internet experiment with a diverse subject pool in Denmark and also find a very high share of conditional cooperators of 70%. Other studies which use different methodologies also find preference heterogeneity in public goods games. See Kurzban and Houser (2005), Bardsley and Moffatt (2007), and Muller, Sefton, Steinberg, and Vesterlund (2008).

why contributions decline is not preference heterogeneity per se but rather many people's tendency to be imperfect conditional cooperators. These people contribute more the more others contribute but nevertheless do not fully match them (their contributions are below the 45-degree line; see Figure 4, left-hand panel). That reason alone is sufficient to explain why the Tragedy of the Commons is almost inevitable. The presence of free riders just speeds up the decline.

### 3.2. Discussion

Numerous lab experiments have shown that the micro-level motivation of many people is conditional cooperation. Field experiments (e.g., Frey and Meier, 2004; Croson and Shang, 2008) and large-scale experiments with more representative samples (Thöni, Tyran, and Wengström, 2009) support this conclusion. There are several psychological reasons why many people are conditional cooperators: the motives of inequity aversion, reciprocity, and guilt aversion introduced in the previous section can explain conditional cooperation. In addition, mere conformism can also lead to conditional cooperation (Carpenter, 2004). Thus, like with gift exchange in the previous example, we expect conditional cooperation to be a robust micro-level motivation.

The previous example illustrated a case where the microstructural conditions were such that the aggregate-level outcomes look the same as under the prediction of self-interest, despite knowing that not everybody is self-interested. In the previous example this occurred because contracts were complete. However, as the present example illustrates, complete contracts are not necessary for a macrobehavioral outcome that resembles one that is predicted if purely self-interested people interact.

There are two important caveats to the bleak prediction of a breakdown of cooperation. Both caveats have to do with the microstructure of interactions. The first one is that in reality people can often choose with whom they interact, whereas in the experiments discussed here the matching into groups occurs at random. One interesting question that follows from the results of Figure 4 (left-hand panel) is whether conditional cooperators who know that all other group members are like-minded cooperators as well could prevent or at least slow down the decline of cooperation. A second caveat is that in the public goods game targeted punishment is not possible. If a conditional cooperator feels duped by his or her less cooperative group members the only possibility to punish is to withdraw contributions. Such a withdrawal also punishes the cooperative group members and is bound

to exacerbate the spiraling down of cooperation. Can targeted punishment prevent the breakdown of cooperation? Numerous experiments suggest yes, but we will show that there are large cross-cultural differences in the ability of punishment to sustain cooperation. In the following two sections we investigate the two caveats in turn.

#### 4. WAYS OUT OF THE TRAGEDY OF THE COMMONS: SORTING?

One microstructural feature behind the experiments investigated in the previous sections is that players are matched randomly. This assumption is often not satisfied in reality because people can choose with whom to interact and this fact alone can change the contribution patterns. We now turn to the question of how sorting can help maintaining cooperation. This question is motivated by the observation made in the previous example; many people are conditional cooperators. The question we ask is whether people who interact with other conditional cooperators are able to maintain high levels of cooperation and can prevent, or at least slow down, the decline of cooperation observed in randomly composed groups. Yet how sensitive is this possibility of maintaining cooperation among like-minded cooperators with regard to the mimicry of pretending to be a cooperator? Free-rider types, which according to the evidence discussed in the previous example comprise a sizeable fraction of the population, might have an incentive to pretend being a cooperator because this way they can lure the cooperators into more cooperation. The cooperator might be aware of this possibility and therefore might be less inclined to cooperate. This shows that sorting of like-minded cooperators might only work if the cooperators can credibly “signal” that they are cooperators. In this section we report a new experiment that tests this idea.

##### 4.1. Experiments and Results

In terms of Figure 1, the microstructural details are as follows. We use the same experimental setup as explained in the previous example, except that in this public goods game group size is three and  $\alpha = 1.8$ . In the following we report results from a baseline public goods game and three treatments with a sorting mechanism. In the treatment *Baseline*, subjects simply play the game described above 10 times in stable groups. Group composition is determined at random and all subjects know this is the case.

The three sorting mechanisms all share one feature of the design developed in Gächter and Thöni (2005). That is, prior to playing the

10 periods public goods game, subjects play a one-shot public goods game with identical parameters. After that, all subjects within a session are ranked according to their contribution in the one-shot public goods game. We refer to this one-shot experiment as the *Ranking* experiment.<sup>7</sup> In the framework of Figure 1, *Ranking* serves as a proxy measurement tool for people's micromotivational attitude to voluntary cooperation.<sup>8</sup> New groups are then formed such that subjects in these groups are as homogeneous as possible with regard to their contribution in the *Ranking* experiment. In particular, the three subjects with the highest contribution in the *Ranking* experiment form the first group. The subjects with the fourth to sixth highest contribution constitute a second group and so on. The last group consists of the three subjects who had chosen the smallest contribution in the *Ranking* experiment.

The three sorting treatments differ only with regard to subjects' information about the matching procedure. In the *Sorted NoInfo* treatment subjects do not have any information about the rematching procedure for the repeated public goods game. Subjects are just told that new groups of three subjects have been formed. Moreover, at the time of the *Ranking* experiment subjects are not yet informed about the existence of the second part of the experiment (where the rematching happens). In the *Sorted Info* treatment all subjects learn the structure of the rematching procedure. In addition, subjects also learn the contribution their new two group members chose in the *Ranking* experiment. However, subjects in *Sorted Info* learn about the second part of the experiment and the rematching procedure only *after* they have chosen their contribution in the *Ranking* experiment. Thus, subjects' decisions in *Ranking* are uninfluenced by the rematching procedure and the subsequent experiment. Hence, *Ranking* measures nonstrategic cooperativeness, because in a one-shot game any positive contribution reflects a cooperative attitude. This feature changes with our third treatment, the *Sorted Strategic* treatment. In this treatment subjects receive full information of the rematching procedure *before*

<sup>7</sup>Instructions are available in Gächter and Thöni (2004). The instructions of the new experiments were adapted accordingly and are available upon request.

<sup>8</sup>*Ranking* is a proxy measurement because unlike in the previous example, not the full contribution vector as a function of others' contribution is measured. The reason is that sorting people into groups as a function of their degree of conditional cooperation is much less straightforward on the basis of a contribution schedule than on the basis of a particular contribution level. Moreover, free rider types clearly have an incentive to contribute nothing in *Ranking*, whereas people who do contribute a positive amount cannot be free rider types. Pessimistic conditional cooperator types might be behaviorally equivalent to free rider types, which is why our measurement is only a proxy tool.

choosing the contribution in the *Ranking* treatment. Apart from this, *Sorted Strategic* is identical to *Sorted Info*.

What do our three main treatments measure? Gächter and Thöni (2005) investigated the *Sorted Info* case and showed that sorting enables cooperative subjects to reach and maintain high levels of cooperation. Our two new treatments expand their approach in two ways.

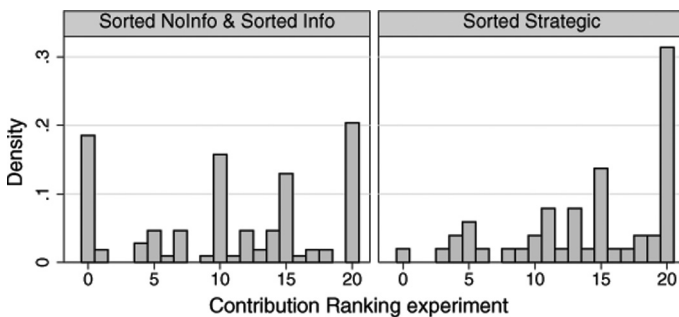
First, by means of our *Sorted Strategic* treatment we address the question whether the functioning of sorting is hampered if the matching procedure is subject to strategic manipulation. To see this, recall that at the time of the *Ranking* experiment subjects in *Sorted Info* do not have prior information about the rematching procedure based on the *Ranking* experiment. Hence, *Ranking* gives us an unbiased measure for a subject's willingness to contribute. In contrast to *Sorted Info*, subjects in *Sorted Strategic* know about the details of rematching when they make their decisions in *Ranking*. Knowledge of the rematching mechanism based on *Ranking* changes the incentive structure of the *Ranking* experiment. Subjects must not only choose a contribution in *Ranking* but also must consider the consequences of their decision for group composition in the second part of the experiment. We conjecture that subjects will choose higher contributions in the *Ranking* experiment of the *Sorted Strategic* treatment compared to the other two treatments. The reason is that free rider types now have an incentive to mimic cooperator types by choosing a high contribution in the *Ranking* experiment in order to increase the chances of being matched with (truly) cooperative subjects later on. If all subjects reason like this, then the *Ranking* experiment loses its signaling value (Frank, 1988, ch. 5). A contribution in *Ranking* may now be strategically biased and therefore not be an indicator of a player's nonstrategic cooperativeness (unlike in *Ranking* in *Sorted Info*, where contributions by design are not strategically biased and therefore credible signals of cooperativeness). We conjecture that in case of the *Sorted Strategic* treatment sorting does not allow high contributors to reach as efficient outcomes as in the *Sorted Info* treatment. The reason is that cooperative subjects might understand the incentives of free rider types to fake being cooperators and therefore may not believe they are interacting with other like-minded cooperators. Moreover, the free rider types may start exploiting the cooperators as the experiment progresses.

Second, we investigate whether, in terms of achieved macro-level contributions, subjects must be aware of the fact that they are sorted with like-minded players or whether it suffices (in the sense of similar contribution levels in *Sorted Info* and *Sorted NoInfo*) to simply match

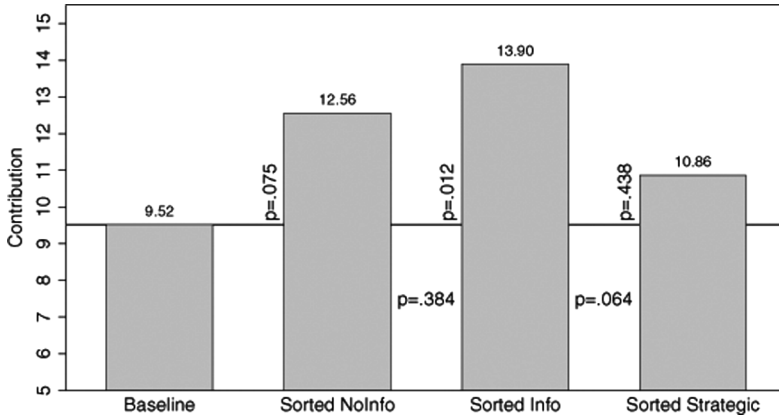
like-minded players together, because like-minded people just match each others' contributions.

We report results from nine sessions with a total of 210 participants (from the University of St. Gallen), which are divided approximately evenly across the four treatments. No subject took part in more than one experiment. In a first step we focus on the results of the *Ranking* experiment of the three main treatments. We investigate whether subjects reacted to the strategic incentive provided in *Sorted Strategic* relative to the other two treatments. Recall that subjects in *Sorted NoInfo* and *Sorted Info* are in exactly the same position when choosing the contribution in the *Ranking* experiment, so we pool the observations of the two treatments. Consistent with our conjecture we observe significantly higher contributions in the *Ranking* experiment of the *Sorted Strategic* treatment relative to the other two treatments (14.2 vs. 10.8,  $z = -2.84$ ,  $p = .005$ , two-sample Wilcoxon rank-sum test). Figure 5 illustrates this finding by showing histograms of the contributions in the *Ranking* experiment. Particularly noteworthy is the observation that in *Sorted Strategic* full contribution was much more frequent compared to the other treatments while full free-riding was almost absent.

In the next step, we address our main research question and compare the three sorting mechanisms with regard to their effect on subsequent macro-level contributions in the repeated public goods game. Figure 6 shows the average contributions during the 10 periods and for all four treatments. Vertical numbers in Figure 6 represent the p-values of two-sample Wilcoxon rank-sum tests comparing each of



**FIGURE 5** Histograms of the contributions in the *Ranking* experiment and for the *Sorted NoInfo* and the *Sorted Info* treatment (left panel) as well as the *Sorted Strategic* treatment (right panel). Data source: Gächter and Thöni (2005), for *Sorted Info*, and new experiments for *Sorted NoInfo* and *Sorted Strategic*.



**FIGURE 6** Average contributions in the ten periods for the four treatments. Numbers in bars indicate bar height;  $p$ -values stem from two-sample Wilcoxon rank-sum test comparing each of the three main treatment to the *Baseline* treatment (vertical  $p$ -values) or two adjacent bars (horizontal  $p$ -values). *Data source:* Gächter and Thöni (2005), for *Baseline* and *Sorted Info*, and new experiments for *Sorted NoInfo* and *Sorted Strategic*.

the three main treatments to the *Baseline* treatment.<sup>9</sup> The horizontal  $p$ -values in Figure 6 depict the significance of the difference between two adjacent bars.

Contributions were lowest in the *Baseline*.<sup>10</sup> In *Sorted NoInfo*, average contributions were weakly significantly higher than in *Baseline*. Thus, simply matching like-minded players has a cooperation-enhancing effect. We observe significantly higher contributions in the treatment *Sorted Info* than in the other two treatments. This finding suggests that in addition to simply matching like-minded

<sup>9</sup>Tests are calculated on the basis of independent group averages. Subjects do only receive information about the contributions of the *Ranking* experiment of their new group members after the rematching procedure. Subjects are not informed about the outcome of the one-shot public goods game of the *Ranking* experiment until the very end of the whole experimental session. Thus, only within rematched groups information spillovers are possible.

<sup>10</sup>In the treatment *Baseline* we combine the results of two treatments: (i) a treatment where subjects simply play the 10 period public goods game with random matching, and (ii) a treatment where subjects first play the *Ranking* experiment and then play ten periods of the public goods game, also with random matching. We use these two treatments to check whether the presence of the *Ranking* treatment has an influence on contributions in the ten period public goods game. It turns out that this is not the case. The treatment averages do not differ according to a two-sample Wilcoxon rank-sum test:  $z = .674$ ;  $p = .501$ .



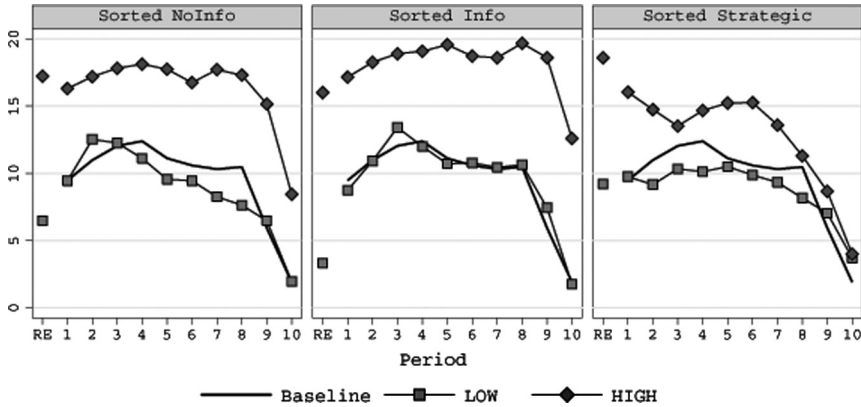
cooperators together (as is done in *Sorted NoInfo*), it matters that players know they are sorted on the basis of a strategically unbiased signal of cooperativeness. Finally, contributions in *Sorted Strategic* are between *Baseline* and the two other sorted treatments. The difference to *Baseline* is insignificant while the difference to *Sorted Info* is weakly significant.

Figure 6 suggests that sorting improves overall contributions. The effect is strongest if subjects are made aware of the rematching procedure *after* having chosen their contribution in the *Ranking* experiment (*Sorted Info*). However, rematching without information seems to improve overall contributions (*Sorted NoInfo*). If subjects know about the rematching mechanism in advance, then contributions are not significantly different from the public goods game with random matching.

In the next step, we look behind this aggregate results and investigate how the different microstructural properties of sorting allow cooperative subjects to maintain high levels of cooperation. For expositional ease we divide our samples in the three main treatments into two categories, each containing half of the groups of the respective treatment.<sup>11</sup> For the categorization we use a median split of the sample according to the groups' average contribution in the *Ranking* experiment. Thus, the category HIGH contains the groups composed of the most cooperative subjects according to the *Ranking* experiment, and LOW contains the below-median cooperative subjects. In Figure 7 we depict the average contributions over the ten periods and for LOW and HIGH in the three main treatments. The leftmost dots represent the average contribution of the group members in the *Ranking* experiment. For comparison we add the average contribution of the *Baseline* treatment to all three graphs (bold line).

Figure 7 shows that both *Sorted NoInfo* and *Sorted Info* produce a sharp distinction between the groups consisting of cooperative subjects and less cooperative subjects according to the *Ranking* experiment. Interestingly, even the groups in LOW manage to reach almost identical contribution levels as observed in the *Baseline* treatment. Thus, the rematching procedure helps cooperative subjects to reach high contribution levels but does not seem to decrease contributions among uncooperative subjects. In the treatment *Sorted Strategic*, the separation between LOW and HIGH is much less pronounced, especially in later periods. Thus, when the *Ranking* experiment does

<sup>11</sup>In the treatment *Sorted Strategic* we have an odd number of groups (17). The category HIGH contains one group more than the category LOW.



**FIGURE 7** Microstructural features of assortative matching and their impact on average cooperation levels of above (HIGH) and below (LOW) median cooperators (according to the *Ranking* experiment (RE)). For ease of comparison each panel contains the average contributions in the *Baseline* treatment. Data source: Gächter and Thöni (2005), for *Baseline* and *Sorted Info*, and new experiments for *Sorted NoInfo* and *Sorted Strategic*.

not reveal a credible signal, the effects of assortative matching are clearly much weaker because those who faked being cooperators start exploiting the true cooperators and the cooperators might be less confident they are among like-minded cooperators as well.

### 4.2. Discussion

In reality, people can often choose with whom to associate, and therefore it is an interesting question how this fact alone shapes cooperation. One of the first studies that looked at this question was Ehrhart and Keser (1999). In their experiments people can choose the group in which they want to be. They find that free riders seek out groups with high cooperation, with the consequence that the cooperators are constantly “on the run” to escape the free riders. Therefore, the decline of cooperation appears inevitable. Our results from the *Strategic Info* treatment are consistent with this result—cooperators cannot be sure to be among like-minded cooperators. An additional reason for the findings of Ehrhart and Keser (1999) might lie in the fact that in their experiment people could only choose the group, not individuals. Coricelli, Fehr, and Fellner (2004) and Page, Putterman, and Unel (2005) show that being able to choose individual group members may be decisive for achieving high contribution levels. In Page

et al.'s experiment, for example, subjects can rank others several times during the experiment, and this ranking determines group composition. This mechanism ensures that the lowest contributors might be removed from the group and it also creates incentives for the free rider types to constantly contribute to the group to prevent being voted out. Our observation from the *Ranking* experiment in *Sorted Strategic* (see Figure 5) shows that free rider types do understand the strategic incentives of pretending to be a cooperator. Our experiment, in comparison to Page et al.'s experiment, also shows that without the continual reinforcement of incentives to behave cooperatively, cooperators soon are unwilling to continue contributing high levels. Finally, similar to our *Sorted NoInfo* treatment, Gunnthorsdottir, Houser, and McCabe (2007) investigate whether simple regrouping of subjects according to their contribution in every round can prevent the decline of cooperation. The idea is that the regrouping of similarly cooperative individuals should provide the cooperators with similar feedback in each round and therefore cooperation should be stabilized. This is exactly what Gunnthorsdottir et al. find. Our experiments show that cooperation is also stabilized (except for an endgame effect) if people are only regrouped once at the beginning.

In summary, the microstructure of interaction and the credibility of others' true cooperativeness are important for convincing the cooperators that they are among like-minded cooperators who among themselves are able to prevent the Tragedy of the Commons. However, if, for whatever reason, no credible signal for cooperativeness exists, cooperation cannot be stabilized through sorting alone. In our last example, we investigate the role of targeted punishment to avoid the Tragedy of the Commons.

## 5. WAYS OUT OF THE TRAGEDY OF THE COMMONS: PUNISHMENT?

All societies use various forms of punishment to sanction antisocial behavior (Coleman, 1990), and it has long been recognized that punishment might solve the free rider problem (e.g., Olson, 1965). Such a demand for punishment might arise in particular in public goods situations and other types of social dilemmas. Some sociologists, such as Coleman (1990), have argued that actions with externalities generate interest among those who experience the externality. This is one condition for the emergence of a demand for a social norm that regulates behavior, and a social norm includes punishment for deviant behavior. Another condition is that a particular action creates similar externalities for a group of people. Thus, people might be willing to

punish free riders. However, punishment is a public good itself—if I punish you do not have to, so I can free ride on your punishment. Therefore, if people are rational and self-regarding, punishment cannot solve the free rider problem, unless the interaction is indefinitely repeated.

Early experiments by Yamagishi (1986) and Ostrom, Walker, and Gardner (1992) showed that the prediction of no punishment is clearly wrong. Many people are indeed willing to expend resources to punish. Fehr and Gächter (2000) and Fehr and Gächter (2002) build on these findings and develop an experimental design that also controls for the strategic incentives to cooperate and punish. Their experiments consist of two stages and the microstructural environment is as follows. The first stage is identical to the baseline public goods experiments described in the previous two sections. After all players have made their contributions, all group members are informed about all contributions in their group. In the second stage they can punish other subjects by reducing their income. Punishment is, however, costly to the punisher himself. Every unit of punishment reduces the income of the punished player by three monetary units and costs the punisher one unit. Fehr and Gächter also vary the microstructure of interactions: in one treatment of Fehr and Gächter (2000), group composition stays constant for the ten periods of the experiment (“Partners”), in another treatment the group composition changes randomly from round to round (“Strangers”). In Fehr and Gächter (2002) group composition changes in every round such that nobody interacts more than once with the same set of group members (“Perfect Strangers”). These microstructural features have a strong impact on macro-level cooperation: in the treatments without punishment average contribution levels are higher among Partners than Strangers or Perfect Strangers. This is also true with punishment. Still, in all experiments punishment is highly effective in increasing and stabilizing cooperation. This is in particular true among Partners where almost full cooperation is achieved. Thus, despite being a “second-order public good,” punishment is very effective in achieving high levels of cooperation.

Subsequent research (see Gächter and Herrmann, 2009, for an overview) has shown that the severity of punishment and its cost have a strong impact on cooperation. If punishment is too expensive or too ineffective, cooperation levels suffer (Egas and Riedl, 2008; Nikiforakis and Normann, 2008). Even the feedback format might matter (Nikiforakis, 2010). Cooperation also suffers if there is uncertainty about others’ possible contributions (Bornstein and Weisel, 2010; Grechenig, Nicklisch, and Thöni, 2010) and if punishment

spirals into “feuds” and “counter-punishment” (Denant-Boemont, Masclet, and Noussair, 2007; Nikiforakis, 2008). Thus, in terms of Figure 1, which macrobehavioral result can be achieved depends on the micro-level details of punishment and the microstructural environment that determines whether interactions are random or repeated.

In the following we continue discussing the experiments of Herrmann et al. (2008a), who also conducted public goods experiments with punishment in 16 comparable subject pools from different societies around the world. The previous examples reported on experiments which were all conducted in one subject pool and hence, in terms of Figure 1, with no variation in the macrosocial environment. In the experiments by Herrmann et al., the micro-level incentives of cooperation and punishment are exactly the same, but the macrosocial environment varies because the experiments were conducted with subject pools from vastly different societies. Thus, their experiments not only allow us to look at how micro-level features of punishment impact on aggregate cooperation levels observed, but also the large cross-cultural variability of their subject pools also allows us to shed light on any link that might exist between micro-level punishment and the macrosocial background of the respective society.

## 5.1. Experiments and Results

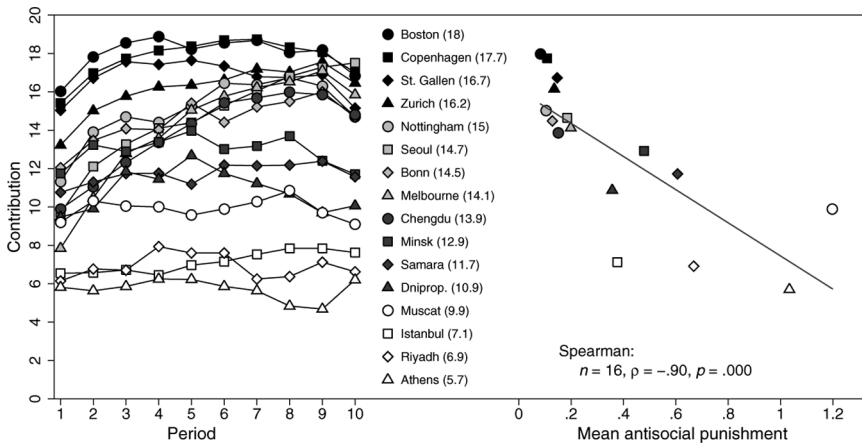
Herrmann et al. (2008a) are interested in the cross-cultural stability of the results observed in previous experiments, which were almost exclusively conducted in a small set of Western countries with little variation in the societal and cultural background. Thus, using the framework of Figure 1, due to a lack of variation the question whether the macrosocial environment has an impact on micromotives and macrobehavior could not be answered. To achieve this variation in the macrosocial environment, Herrmann et al. ran their experiments in countries which differed strongly according to measures developed by other social scientists: trust and civic norms of cooperation (World Values Survey), the rule of law and GDP per capita, democracy, cultural dimensions, and value orientations. To be able to draw inferences of the importance of the macrosocial environment, the microstructural environment (i.e., all rules and procedures of the experiment) was kept constant.

Two features of punishment are striking. First, people punish free riders (defined as people who contribute *less* than the punisher) rather similarly everywhere. Second, there are large differences with regard to “antisocial punishment” defined as punishment of people who contribute *the same or more* as the punisher. In some subject pools

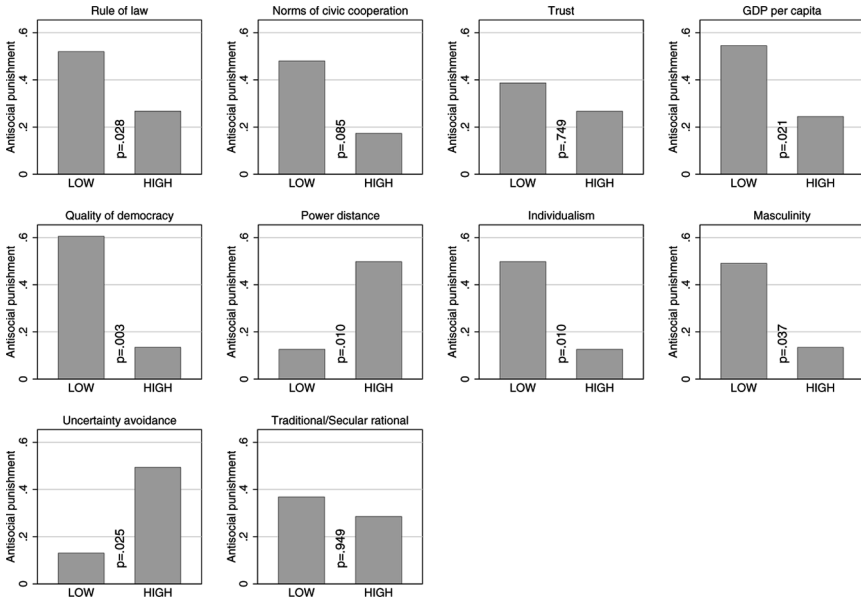
antisocial punishment is almost negligible, whereas in others it is substantial. On the  $x$ -axis of the right panel in Figure 8 we denote the average number of punishment points subjects assigned to other subjects with equal or higher contribution than themselves. Thus, the methodology of holding the microstructural environment constant and varying the macrosocial environment revealed stark differences in behavior.

Figure 8 shows the macrobehavioral outcome in terms of the average cooperation level per subject pool. The left panel shows that punishment stabilizes cooperation everywhere but at vastly different levels. The right panel relates cooperation levels to the extent of antisocial punishment in a subject pool. It turns out that mean contributions are strongly negatively related to antisocial punishment. A further statistical analysis shows that cooperation increases in punishment of free riding and decreases in antisocial punishment. Thus, if the micromotive of antisocial punishment is prevalent, punishment does not solve the Tragedy of the Commons.

The large variation in punishment behavior across subject pools and the large number of independent groups who took part in these experiments ( $n = 280$  groups) allows us to link the microfeatures of punishment to the macro-outcome of average cooperation levels. Yet, the large variation in cultural background of these subject pools



**FIGURE 8** Average contributions over time (left panel) and the relationship of antisocial punishment and average cooperation levels (right panel). Mean antisocial punishment measures the number of punishment points subjects assign to other subjects with equal or higher contribution than themselves. *Source:* Adapted from Herrmann et al. (2008a).



**FIGURE 9** The link between antisocial punishment and the macrosocial environment according to ten different indicators. For exact definitions of the variables and their data sources see Herrmann et al. (2008b). We applied a median split of the scores of the relevant variables to group subject pools into LOW and HIGH groups, respectively. *P*-values refer to Mann-Whitney tests with subject pool averages as independent observations. *Source*: Own analysis based on data from Herrmann et al. (2008a).

permits investigating a relationship the previous experiments discussed here could not address: Is there a link between the societal background as measured by various variables developed to measure cross-cultural differences and the microlevel patterns of punishment?

We look at this question by dividing the subject pools for each of the cross-cultural variables introduced above into two groups according to a median split (in our sample of societies). Thus, we group the subject pools in two categories: LOW if the score of the respective variable is below the median in the sample and HIGH if it is above the median in the sample. We then calculate the mean antisocial punishment for LOW and HIGH. Figure 9 depicts the results. For example, subject pools that come from societies with below-median scores on the rule of law indicator exhibit on average almost twice as much antisocial punishment than subject pools from societies with above-median strength of the rule of law. Of the 10 indicators depicted in Figure 9, seven show

a significant difference between the LOW and the HIGH group and one returns a weakly significant difference. If we perform the same analysis for free rider punishment we find no difference at all. Thus, punishment of free riding behavior does not seem to be influenced by the cultural background, whereas antisocial punishment seems to be strongly shaped by it (see also Gächter, Herrmann, and Thöni, 2010, for a further analysis of this point).

## 5.2. Discussion

The second-order free rider problem behind punishment that theoretically undermines the possibility of punishment to mitigate the free rider problem is empirically unimportant.<sup>12</sup> This is true despite the fact that many people are second-order free riders. Many people actually do punish, however, and in those subject pools in which the punishers direct their sanctions at the free riders only, cooperation flourishes. If the fact that free riding will get punished is sufficiently established, punishment can work as a mere threat and the actual costs of punishment are low (Gächter, Renner, and Sefton, 2008). Moreover, recent experiments show that a little punishment can help rewarding strategies to flourish (Ule, Schram, Riedl, and Cason, 2009) and that communication (combined with punishment) can be very effective (Janssen, Holahan, Lee, and Ostrom, 2010).

There are several reasons people punish free riders, including negative emotions, equity concerns, or strong negative reciprocity (for a discussion see Gächter and Herrmann, 2009). Again, this multitude of possible micromotives for punishment suggests that punishment is a robust phenomenon. The question why people punish antisocially is open for future research (Thöni, 2010). The existing evidence suggests that revenge is a likely motive (Herrmann et al., 2008a), but other motives are most likely relevant as well because revenge cannot explain antisocial punishment in many cases.

Irrespective of the psychological sources of punishment, the experiments of Herrmann et al. (2008a) clearly demonstrate that microlevel motivations of punishment have strong macrobehavioral consequences with regard to the extent to which the Tragedy of the Commons can be mitigated. This finding complements theoretical and empirical insights about the importance of the microstructural environment for

<sup>12</sup>There are only a few studies that have utilized non-student subject pools to study punishment behavior (Egas and Riedl, 2008; Gächter and Herrmann, in press). These studies find limited evidence for the importance of socio-demographics for punishment behavior.



sustaining cooperation (Ostrom, 1990; Ostrom et al., 1992; Carpenter, 2007; Egas and Riedl, 2008; Nikiforakis and Normann, 2008; Ertan, Page, and Putterman, 2009).

## 6. CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

Understanding how individual decisions shape aggregate outcomes in a particular decision environment is of fundamental importance in the social sciences. In addition to (game-theoretic) modeling and simulations, we believe that experiments have a useful complementary role to play because they allow us to observe real behavior but in theoretically informed and controlled decision environments. Experiments have served the natural sciences well, and some important differences notwithstanding, they are also an invaluable tool for understanding social reality and to inform new theories on human behavior. The experiments discussed in this article should be seen as “exhibits” (Sugden, 2005) for this approach, whose applicability extends beyond the topic of voluntary cooperation.

What is the overall message from our four exhibits? We can again use Figure 1 to structure this discussion.

1. In terms of measured *micromotives* our results reinforce conclusions reached in earlier related papers (e.g., Fehr and Gintis, 2007): the assumption that most people are self-regarding is strongly refuted in favor of strong positive and negative reciprocity as motives that are empirically mostly more important than selfishness. However, a substantial number of people are best characterized as selfish. Hence, heterogeneity in pro-social motivations is an important empirical fact found in all examples (and highlighted in the first three examples).
2. In our examples, the consequences of heterogeneous micromotives for macrobehavioral outcomes depend also on *microstructures*. The reason this is not trivial is that all experiments discussed here were designed such that under selfishness different microstructure would not change the theoretical prediction of the observed macrobehavioral outcomes. The experiments highlight that, depending on the details of the microstructure, the macrobehavioral outcome either resembles one that is predicted by universal selfishness (example 1 under complete contracts and example 2) or one where the macrobehavioral outcome is shaped by the strong reciprocators (all examples contained this case).

3. As demonstrated in our fourth example, the *macrosocial environment* can also exert a strong influence on micromotives and, together with the microstructure, also on the macrobehavioral outcome.

In terms of the theoretical implications of our results we believe they support modeling strategies based on a rational choice approach. The results of the experiments discussed here only firmly reject the assumption that people are selfish. However, selfishness is *not* an ingredient of rational choice approaches; therefore, the fact that many people are not selfish cannot be taken as a refutation of a rational choice approach (see Gintis, 2009, for a comprehensive defense of this viewpoint). In fact, the results of all four examples are consistent with rational choice approaches, if we assume that a majority of people is not selfish, an assumption vindicated by the empirical facts. Using formal models and applying a rational choice perspective has also considerably enhanced our understanding of the role of inequity aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), reciprocity (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006), and altruism (Cox et al., 2008) for various important social outcomes.

We see three directions for future research:

1. We start again with *micromotives*. There can be no doubt that people's micromotives are heterogeneous. Research has started to understand how preferences are distributed across different populations of interest, but clearly more research is needed (Henrich, Heine, and Norenzayan, 2010). Field experiments are an important complementary tool in this regard. In our view a particularly fruitful approach is to integrate experiments into social surveys (household panels) because this allows combining the advantages of survey-based research with those of behavioral research (e.g., Fehr, Fischbacher, Von Rosenblatt, Schupp, and Wagner, 2002; Bellemare and Kröger, 2007; Thöni et al., 2009). This line of research allows understanding how sociodemographics influence micromotives. We also need to better understand the (neuro-) psychological underpinnings of people's micromotives. Some encouraging results have already been collected (Glimcher, Camerer, Fehr, & Poldrack, 2009). The observation of heterogeneous micromotives also raises the question whether there are stable interindividual differences or whether people are sometimes motivated by strong reciprocity and sometimes by selfishness. The micromotives we have looked at in this study are primarily selfishness and strong positive and negative reciprocity.

However, the phenomenon of antisocial punishment also shows that a somewhat darker side of human motivations exists as well and this sort of spitefulness needs more research.

2. Heterogeneity in micromotives and the fact that many people are strong reciprocators also raises important further questions on interaction effects of *microstructures* with micromotives. For instance, incentive systems that stress people's self-interest may crowd out pro-social motivations (e.g., Bowles, 2008). Therefore, understanding interaction effects of various microstructures with people's (nonselfish) micromotives is of great social and economic importance and may lead to different results than similar such investigations assuming self-regarding actors.
3. With regard to the impact of the *macrosocial environment* on micromotives, we believe research is still in its infancy. Great progress has been made by anthropologists who study how the macrosocial environment influences norms of fairness across various small-scale societies (e.g., Henrich et al., 2010). Much less is understood about large-scale modern societies. Herrmann et al. (2008a) is surely only a beginning, and the research agenda extends beyond the specific issues discussed in this article. For example, are there also differences in risk preferences or inter-temporal preferences across societies? Moreover, macro-social differences can also exist within societies, not only between societies (see, e.g., Gächter and Herrmann, in press, who look at differences between urban and rural populations in Russia).

We close by reiterating one main message of this article here: Laboratory experiments are an apt tool to complement theoretical and empirical investigations into the micro-macro link. We hope to have shown the power of lab experiments for the case of voluntary cooperation. Yet, the power of lab experiments surely extends beyond this topic.

## REFERENCES

- Adams, J. S. (1965). Inequity in social exchange. In J. S. Adams (Ed.), *Advances in Experimental Psychology* (pp. 267–299). New York, NY: Academic Press.
- Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *Quarterly Journal of Economics*, 97, 543–569.
- Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., & Sugden, R. (2010). *Experimental Economics: Rethinking the Rules*. Princeton, NJ: Princeton University Press.
- Bardsley, N. & Moffatt, P. G. (2007). The experimetrics of public goods: inferring motivations from contributions. *Theory and Decision*, 62, 161–193.

- Bellemare, C. & Kröger, S. (2007). On representative social capital. *European Economic Review*, 51, 183–202.
- Benz, M. & Meier, S. (2008). Do people behave in experiments as in the field?—Evidence from donations. *Experimental Economics*, 11, 268–281.
- Bolton, G. E. & Ockenfels, A. (2000). ERC: a theory of equity, reciprocity, and competition. *American Economic Review*, 90, 166–193.
- Bornstein, G. & Weisel, O. (2010). Punishment, cooperation, and cheater detection in “noisy” social exchange. *Games*, 1(1), 18–33.
- Bowles, S. (2003). *Microeconomics: Behavior, Institutions, and Evolution*. Princeton, NJ: Princeton University Press.
- Bowles, S. (2008). Policies designed for self-interested citizens may undermine “the moral sentiments”: evidence from economic experiments. *Science*, 320(5859), 1605–1609.
- Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 3531–3535.
- Boyd, R., Gintis, H., & Bowles, S. (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science*, 328(5978), 617–620.
- Buskens, V. & Raub, W. (2008). *Rational Choice Research on Social Dilemmas: Embeddedness Effects on Trust* (ISCORE Paper 200). Utrecht, The Netherlands: Utrecht University.
- Carpenter, J. (2004). When in Rome: conformity and the provision of public goods. *Journal of Socio-Economics*, 33, 395–408.
- Carpenter, J. (2007). Punishing free-riders: how group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior*, 60, 31–51.
- Carpenter, J., Harrison, G., & List, J. (2005). *Field Experiments in Economics*. Greenwich, CT: JAI Press.
- Charness, G. & Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74, 1579–1601.
- Coleman, J. (1990). *Foundations of Social Theory*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Colman, A. M. (1999). *Game Theory and Its Applications in the Social and Biological Sciences*. London, UK: Routledge.
- Coricelli, G., Fehr, D., & Fellner, G. (2004). Partner selection in public goods experiments. *Journal of Conflict Resolution*, 48, 356–378.
- Cox, J. C., Friedman, D., & Sadiraj, V. (2008). Revealed altruism. *Econometrica*, 76, 31–69.
- Croson, R. (2007). Theories of commitment, altruism and reciprocity: evidence from linear public goods games. *Economic Inquiry*, 45, 199–216.
- Croson, R. & Gächter, S. (2010). The science of experimental economics. *Journal of Economic Behavior & Organization*, 73, 122–131.
- Croson, R. & Shang, J. (2008). The impact of downward social information on contribution decisions. *Experimental Economics*, 11, 221–233.
- Davis, D. D. & Holt, C. A. (1993). *Experimental Economics*. Princeton, NJ: Princeton University Press.
- Denant-Boemont, L., Masclet, D., & Noussair, C. N. (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory*, 33(1), 145–167.
- Diekmann, A. (2004). The power of reciprocity: fairness, reciprocity, and stakes in variants of the dictator game. *Journal of Conflict Resolution*, 48, 487–505.

- Dufwenberg, M. & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, 47, 268–298.
- Egas, M. & Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B: Biological Sciences*, 275(1637), 871–878.
- Ehrhart, K.-M. & Keser, C. (1999). *Mobility and Cooperation: On the Run* (Working Paper No. 99s-24). Montreal, Quebec, Canada: CIRANO.
- Ellingsen, T. & Johannesson, M. (2008). Pride and prejudice: the human side of incentive theory. *American Economic Review*, 98, 990–1008.
- Ertan, A., Page, T., & Putterman, L. (2009). Who to punish? Individual decisions and majority rule in mitigating the free rider problem. *European Economic Review*, 53, 495–511.
- Falk, A. & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54, 293–315.
- Falk, A., Gächter, S., & Kovacs, J. (1999). Intrinsic motivation and extrinsic incentives in a repeated game with incomplete contracts. *Journal of Economic Psychology*, 20, 251–284.
- Falk, A. & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952), 535–538.
- Fehr, E., Fischbacher, U., Von Rosenblatt, B., Schupp, J., & Wagner, G. G. (2002). A nationwide laboratory. Examining trust and trustworthiness by integrating behavioral experiments into representative surveys. *Schmoller's Jahrbuch*, 122, 519–542.
- Fehr, E. & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90, 980–994.
- Fehr, E. & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137–140.
- Fehr, E. & Gintis, H. (2007). Human motivation and social cooperation: experimental and analytical foundations. *Annual Review of Sociology*, 33, 43–64.
- Fehr, E., Goette, L., & Zehnder, C. (2009). A behavioral account of the labor market: the role of fairness concerns. *Annual Review of Economics*, 1, 355–384.
- Fehr, E., Kirchler, E., Weichbold, A., & Gächter, S. (1998). When social norms overpower competition: gift exchange in experimental labor markets. *Journal of Labor Economics*, 16, 324–351.
- Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. *Quarterly Journal of Economics*, 108, 437–459.
- Fehr, E. & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114, 817–868.
- Fischbacher, U. & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public good experiments. *American Economic Review*, 100, 541–556.
- Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71, 397–404.
- Frank, R. H. (1988). *Passion Within Reason: The Strategic Role of the Emotions*. New York, NY: W.W. Norton & Company.
- Frey, B. S. & Meier, S. (2004). Social comparisons and pro-social behavior. Testing “conditional cooperation” in a field experiment. *American Economic Review*, 94, 1717–1722.
- Gächter, S. & Herrmann, B. (2009). Reciprocity, culture, and human cooperation: previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1518), 791–806.

- Gächter, S. & Herrmann, B. (in press). The limits of self-governance when cooperators get punished—experimental evidence from urban and rural Russia. *European Economic Review*. doi:10.1016/j.euroecorev.2010.04.003
- Gächter, S., Herrmann, B., & Thöni, C. (2010). Culture and cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356, 2651–2661. doi:10.1098/rstb.2010.0135
- Gächter, S., Renner, E., & Sefton, M. (2008). The long-run benefits of punishment. *Science*, 322(5907), 1510.
- Gächter, S. & Thöni, C. (2004). *Social Learning and Voluntary Cooperation Among Like-Minded People* (Discussion Paper No. 2004-12). St. Gallen, Switzerland: University of St. Gallen.
- Gächter, S. & Thöni, C. (2005). Social learning and voluntary cooperation among like-minded people. *Journal of the European Economic Association*, 3(2–3), 303–314.
- Gächter, S. & Thöni, C. (2010). Social comparison and performance: experimental evidence on the fair wage–effort hypothesis. *Journal of Economic Behavior & Organization*, 76, 531–543.
- Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206, 169–179.
- Gintis, H. (2009). *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton, NJ: Princeton University Press.
- Glimcher, P. W., Camerer, C. F., Fehr, E., & Poldrack, R. A. (2009). *Neuroeconomics. Decision Making and the Brain*. Amsterdam, The Netherlands: Elsevier.
- Gouldner, A. W. (1960). The norm of reciprocity: a preliminary statement. *American Sociological Review*, 25, 161–178.
- Grechenig, C., Nicklisch, A., & Thöni, C. (2010). Punishment despite reasonable doubt—A public goods experiment with sanctions under uncertainty. *Journal of Empirical Legal Studies*, 7, 847–867.
- Gunthorsdottir, A., Houser, D., & McCabe, K. (2007). Disposition, history and contributions in public goods experiments. *Journal of Behavior and Organization*, 62, 304–315.
- Hardin, G. (1968). The tragedy of the commons. *Science*, 162(3859), 1243–1248.
- Hechter, M. & Horne, C. (2009). *Theories of Social Order: A Reader* (2nd ed.). Stanford, CA: Stanford University Press.
- Hechter, M. & Kanazawa, S. (1997). Sociological rational choice theory. *Annual Review of Sociology*, 23, 191–214.
- Helbing, D. & Yu, W. J. (2009). The outbreak of cooperation among success-driven individuals under noisy conditions. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 3680–3685.
- Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., ... Ziker, J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science*, 327(5972), 1480–1484.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33, 61–135.
- Herrmann, B. & Thöni, C. (2009). Measuring conditional cooperation: a replication study in Russia. *Experimental Economics*, 12(1), 87–92.
- Herrmann, B., Thöni, C., & Gächter, S. (2008a). Antisocial punishment across societies. *Science*, 319(5868), 1362–1367.
- Herrmann, B., Thöni, C., & Gächter, S. (2008b). Supporting online material for antisocial punishment across societies. *Science*, 319(5868). Retrieved from <http://www.sciencemag.org/cgi/data/319/5868/1362/DC/1>

- Janssen, M. A., Holahan, R., Lee, A., & Ostrom, E. (2010). Lab experiments for the study of social-ecological systems. *Science*, *328*(5978), 613–617.
- Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *Science*, *322*(5908), 1681–1685.
- Kelley, H. & Stahelski, A. (1970). Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology*, *16*, 190–219.
- Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J., & Sutter, M. (2008). Conditional cooperation on three continents. *Economics Letters*, *101*, 175–178.
- Kollock, P. (1998). Social dilemmas: the anatomy of cooperation. *Annual Review of Sociology*, *24*, 183–214.
- Kurzban, R. & Houser, D. (2005). Experiments investigating cooperative types in humans: a complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 1803–1807.
- Levitt, S. D. & List, J. A. (2008). Homo economicus evolves. *Science*, *319*(5865), 909–910.
- Loewenstein, G., Thompson, L., & Bazerman, M. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology*, *57*, 426–441.
- Macy, M. W. & Willer, R. (2002). From factors to actors: computational sociology and agent-based modeling. *Annual Review of Sociology*, *28*, 143–166.
- Muller, L., Sefton, M., Steinberg, R., & Vesterlund, L. (2008). Strategic behavior and learning in repeated voluntary-contribution experiments. *Journal of Economic Behavior & Organization*, *67*, 782–793.
- 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. & Normann, H. (2008). A comparative statics analysis of punishment in public goods experiments. *Experimental Economics*, *11*, 358–369.
- Nikiforakis, N. (2010). Feedback, punishment and cooperation in public good experiments. *Games and Economic Behavior*, *68*, 689–702.
- Olson, M. (1965). *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.
- Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, UK: Cambridge University Press.
- Ostrom, E. (2005). *Understanding Institutional Diversity*. Princeton, NJ: Princeton University Press.
- Ostrom, E., Walker, J. M., & Gardner, R. (1992). Covenants with and without a sword—Self-governance is possible. *American Political Science Review*, *86*, 404–417.
- Page, T., Putterman, L., & Unel, B. (2005). Voluntary association in public goods experiments: reciprocity, mimicry, and efficiency. *Economic Journal*, *115*(506), 1032–1052.
- Rabin, M. (1993). Incorporating fairness into game-theory and economics. *American Economic Review*, *83*, 1281–1302.
- Rapoport, A. & Chammah, A. M. (1965). *Prisoners' Dilemma: A Study in Conflict and Cooperation*. Ann Arbor, MI: The University of Michigan Press.
- Rosser, Jr., J. B. & Eckel, C. (2010). Introduction to JEBO special issue on “issues in the methodology of experimental economics.” *Journal of Economic Behavior & Organization*, *73*, 1–2.
- Sally, D. (1995). Conversation and cooperation in social dilemmas: a meta-analysis of experiments from 1958 to 1992. *Rationality and Society*, *7*(1), 58–92.
- Schelling, T. (1978). *Micromotives and Macrobehavior*. New York, NY: W. W. Norton.
- Sigmund, K. (2010). *The Calculus of Selfishness*. Princeton, NJ: Princeton University Press.

- Smith, V. L. (1982). Microeconomic systems as an experimental science. *American Economic Review*, 72, 923–955.
- Sugden, R. (2005). Experiments as exhibits and experiments as tests. *Journal of Economic Methodology*, 12, 291–302.
- Thöni, C. (2010). *Inequality Aversion and Antisocial Punishment* (Discussion Paper). St. Gallen, Switzerland: University of St. Gallen.
- Thöni, C., Tyran, J.-R., & Wengström, E. (2009). Microfoundations of social capital (Discussion Paper No. 09-24). Copenhagen, Denmark: University of Copenhagen, Department of Economics.
- Ule, A., Schram, A., Riedl, A., & Cason, T. N. (2009). Indirect punishment and generosity toward strangers. *Science*, 326(5960), 1701–1704.
- Vieth, M. (2009). *Commitments and Reciprocity. Experimental Studies on Obligation, Indignation, and Self-Consistency*. Utrecht, The Netherlands: Universiteit Utrecht.
- Webster, M. J. & Sell, J. (2007). *Laboratory Experiments in the Social Sciences*. Amsterdam, The Netherlands: Elsevier.
- Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51, 110–116.