PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2009) **364**, 791–806
doi:10.1098/rstb.2008.0275
Published online 12 December 2008

Review

Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment

Simon Gächter[1,2,3,*] **and Benedikt Herrmann**[1]

[1]*Centre of Decision Research and Experimental Economics, School of Economics, University of Nottingham, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, UK*
[2]*CESifo, Poschingerstrasse 5, 81679 Munich, Germany*
[3]*IZA, Institute for the Study of Labor, Schaumburg-Lippe-Strasse 5-9, 53113 Bonn, Germany*

Understanding the proximate and ultimate sources of human cooperation is a fundamental issue in all behavioural sciences. In this paper, we review the experimental evidence on how people solve cooperation problems. Existing studies show without doubt that direct and indirect reciprocity are important determinants of successful cooperation. We also discuss the insights from a large literature on the role of peer punishment in sustaining cooperation. The experiments demonstrate that many people are 'strong reciprocators' who are willing to cooperate and punish others even if there are no gains from future cooperation or any other reputational gains. We document this in new one-shot experiments, which we conducted in four cities in Russia and Switzerland. Our cross-cultural approach allows us furthermore to investigate how the cultural background influences strong reciprocity. Our results show that culture has a strong influence on positive and in especially strong negative reciprocity. In particular, we find large cross-cultural differences in 'antisocial punishment' of pro-social cooperators. Further cross-cultural research and experiments involving different socio-demographic groups document that the antisocial punishment is much more widespread than previously assumed. Understanding antisocial punishment is an important task for future research because antisocial punishment is a strong inhibitor of cooperation.

Keywords: human cooperation; strong reciprocity; public goods experiments; culture; antisocial punishment

1. INTRODUCTION

Many important collective problems that human decision makers face are characterized by a conflict of interest between individual and group benefit. The 'tragedy of the commons' (Hardin 1968) is probably the best known example. Each individual farmer has an incentive to put as many cattle on the common meadow as possible. The tragic consequence may be over-grazing from which all farmers suffer. Collectively, all farmers would be better off if they were able to constrain the number of cattle that grazes on the commons. Yet, each individual farmer is better off by letting their cattle graze. Collective welfare is jeopardized by individual greed in such diverse areas as warfare; cooperative hunting and foraging; environmental protection; tax compliance; voting; the participation in collective actions such as demonstrations, strikes, embargoes and consumer boycotts; the voluntary provision of public goods; donations to charities; teamwork; collusion between firms; and so on. However, despite this bleak prediction, humans often manage to avoid the tragedy of the commons and achieve high levels of cooperation. This holds for hunter-gatherer societies to complex modern nation states, which would not exist without large-scale cooperation. Thus, understanding cooperation is an important challenge for all social sciences but also for evolutionary biology, because it needs to explain how natural and cultural evolution can lead to cooperation (Hammerstein 2003; Gardner & West 2004; Henrich & Henrich 2007; West *et al.* 2007).

This paper reviews existing evidence and presents novel cross-cultural results from systematic experimental investigations on how people solve cooperation problems. We believe that sound empirical knowledge is an important input for the development of proximate and ultimate theories of cooperation. Laboratory experiments are probably the best tool for studying cooperation empirically. The reason is that in the field many factors are operative at the same time. The laboratory allows for a degree of control that is often not feasible in the field.[1,2]

* Author and address for correspondence: Centre of Decision Research and Experimental Economics, School of Economics, University of Nottingham, Sir Clive Granger Building, University Park, Nottingham, NG7 2RD, UK (simon.gaechter@nottingham.ac.uk).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rstb.2008.0275 or via http://journals.royalsociety.org.

One contribution of 11 to a Theme Issue 'Group decision making in humans and animals'.

In particular, experiments are helpful for separating out explanations why people cooperate. According to some important proximate theories as developed in the social sciences, in particular economics, and ultimate theories as developed in evolutionary biology, people cooperate only if it is in their (long-term) self-interest. For instance, if the interaction is among genetic relatives ('kin selection', Hamilton 1964) or if it is repeated and/or if one's reputation is at stake ('direct reciprocity' and 'indirect reciprocity', respectively), people might have a selfish incentive to cooperate (Trivers 1971; Axelrod & Hamilton 1981; Fudenberg & Maskin 1986; Panchanathan & Boyd 2004; Nowak & Sigmund 2005; Lehmann & Keller 2006; Nowak 2006). Experimental approaches allow the researcher to control, by way of experimental design, the extent to which direct and indirect reciprocity are possible.

This paper presents attempts to study with the help of controlled laboratory experiments some important factors that influence an important aspect of human collective decision making: cooperation behaviour. Our focus is on cooperation because this has been a particularly active research field in the behavioural sciences recently (both theoretically and experimentally), with a large potential for cross-disciplinary research (Hammerstein 2003; Hammerstein & Hagen 2005; Fehr & Camerer 2007; Sigmund 2007).[3]

Our paper is structured as follows. In §2, we introduce our tool of investigation—the public goods game. In §3, we review the evidence that shows that both repeated interaction and possibilities for reputation formation are important determinants for people's cooperation behaviour. However, there is also substantial cooperation in anonymous one-shot games, where neither strategic reciprocity nor reputation can matter. In experiments in which people have the possibility to punish their group members at their own cost after having seen how much the other group members contributed, it turned out that the punishment of freeloaders is an important factor to explain cooperation in both one-shot and repeated interactions. Cooperation in one-shot games is an evidence for 'strong reciprocity' (Gintis 2000; Fehr *et al.* 2002*a*; Fehr & Fischbacher 2003; Carpenter *et al.* in press). Strong reciprocators incur personal costs to punish and reward others even when this behaviour cannot be justified by kinship, reciprocal altruism or reputational concerns. Thus, strong reciprocity presents a particular challenge to evolutionary theories of cooperation and has therefore been an important topic for research in the last few years. We will provide novel evidence for strong reciprocity in §4.

Section 4 is the core of our paper because we show that people cooperate and punish in one-shot games without any repetition. The experiment we report in this section also shows that there exists a substantial cultural influence on strong reciprocity. Section 5 follows up on the findings from §4 by briefly reviewing a cross-cultural experiment conducted in 16 participant pools around the globe (Herrmann *et al.* 2008). This experiment demonstrates that cooperation and punishment are substantially shaped by the cultural background across a range of diverse societies.

Most experiments on strong reciprocity were conducted with students as participants, which raises the question of how general the observations on strong reciprocity are across different socio-economic groups. Section 6 reviews some recent findings on this question. Section 7 provides concluding remarks.

2. THE PUBLIC GOODS GAME

Many human cooperation problems—from hunter-gatherer societies to modern societies—often involve large numbers of individuals. The 'public goods game' is a suitable research tool for studying such n-person cooperation problems.[4] In this game, each of n-group members receives an endowment of, say, 20 tokens. Participants have to decide how many tokens to keep for themselves and how many to contribute to a group project, which collectively earns $n\alpha > 1$ for each token invested. Each group member earns α tokens (where $0 < \alpha < 1$) for each token invested in the project, regardless of whether he or she contributed any. Since the cost of contributing one token to the project is exactly one token while the individual return on that token is only $\alpha < 1$ tokens, keeping all one's own tokens is always in any participant's material self-interest—irrespective of how much the other group members contribute. Yet, if, for example, in a group of four and $\alpha = 0.5$ each group member retains all of his or her tokens then there are no earnings to be shared; on the other hand, each member would earn $0.5 \times 80 = 40$ tokens if each of them invests their entire 20 token endowment.

The public goods game epitomizes the tension between collective welfare and individual incentives in a simple and stark way because selfish rationality implies full 'free riding' (i.e. zero contributions), whereas collective welfare is maximized if every player makes maximal contributions. Owing to its simplicity the public goods game has been used to answer questions about how various institutional parameters, such as group size (n), the marginal gains from cooperation (α), the strategic nature of interaction structures (one-shot versus repeated interaction) and possibilities for multilateral peer punishment influence cooperation. The public goods game is also a prototypical game to study pro-social behaviour in a group context (Camerer & Fehr 2004). We discuss the most important findings in §3.

3. FACTORS THAT INFLUENCE COOPERATION

Under the assumption that agents are rational and want to maximize their monetary pay-off theory predicts that people will not contribute to the public good. However, numerous experiments have falsified this prediction—there exists substantial cooperation in a variety of set-ups.[5] Six sets of results are particularly noteworthy in the light of existing proximate and ultimate theories of cooperation.

(i) Contributions are higher, the higher the marginal gains from contributing (i.e. α) are (Isaac & Walker 1988*b*; Brandts & Schram 2001; Goeree *et al.* 2002; Zelmer 2003; Carpenter 2007*b*). This is interesting because

from the viewpoint of (selfishly) rational decision making the prediction of full free riding in the public goods game described above does not depend on α, as long as $\alpha < 1$. However, this result shows that people apparently find it easier to contribute to the public good the higher the marginal gains from cooperation are (Anderson *et al.* 1998).

(ii) Larger groups do not cooperate significantly less than smaller groups (Marwell & Ames 1979; Isaac & Walker 1988*b*; Isaac *et al.* 1994; Zelmer 2003; Carpenter 2007*b*; Cardenas & Jaramillo 2007). This finding goes against conventional wisdom that maintaining cooperation should be easier in smaller groups (Olson 1965). One explanation might be that people are heterogeneous with respect to their willingness to cooperate (more on this below). Some are 'free riders' and others are 'conditional cooperators' who are willing to cooperate provided others cooperate as well. Larger groups may have more free riders than small groups, but they possibly also have more cooperators. Group size *per se* is therefore not decisive.

(iii) Playing the public goods game repeatedly with the same group members often leads to higher contributions than playing it one shot and with randomly changing group members (Croson 1996; Sonnemans *et al.* 1999; Fehr & Gächter 2000; Keser & van Winden 2000).[6] This finding (and related ones from indefinitely repeated prisoners' dilemma games (Dal Bo 2005)) is consistent with ultimate and proximate arguments that repeated interactions offer strategic reasons to cooperate (Trivers 1971; Axelrod & Hamilton 1981; Kreps *et al.* 1982; Fudenberg & Maskin 1986). The significance of the finding that cooperation is typically higher in repeated games than one-shot games, and similar findings from related cooperation experiments (e.g. Falk *et al.* 1999; Engelmann & Fischbacher 2002; Gächter & Falk 2002; Cochard *et al.* 2004) is that people are able to distinguish situations that require strategic cooperation from those that do not (Fehr & Fischbacher 2003).

(iv) Experiments under non-anonymity, where participants could identify the individual behind a particular contribution, increased contributions relative to an anonymity benchmark (Gächter & Fehr 1999; Andreoni & Petrie 2004; Rege & Telle 2004). People even contribute more to public goods if they are exposed to subconsciously activated cues of being observed (Bateson *et al.* 2006; Burnham & Hare 2007).[7] This evidence is consistent with 'reputation effects' noted in several decision tasks involving altruistic behaviour (Haley & Fessler 2005; Milinski & Rockenbach 2007). People might care for a favourable reputation because this is evolutionarily advantageous according to the models of indirect reciprocity (Nowak & Sigmund 2005), where people are more likely to receive help if they have helped others in the past and therefore have a favourable 'image score'. The experimental evidence is consistent with such a mechanism (Engelmann & Fischbacher 2002; Milinski *et al.* 2002; Semmann *et al.* 2005; Seinen & Schram 2006).

(v) Communication also greatly facilitates cooperation and helps in preventing its breakdown (Dawes *et al.* 1977; Isaac & Walker 1988*a*; Ostrom *et al.* 1992; Sally 1995; Brosig *et al.* 2003; Bochet *et al.* 2006). Similarly, intergenerational advice, if common knowledge, can also sustain high levels of cooperation (Chaudhuri *et al.* 2006). Communication is interesting because it is an important human capacity that can often be fruitfully employed in smaller groups. There are many behavioural reasons why communication is effective: communication might help the cooperators to coordinate on high levels and it might involve social pressure and mutual promises which would induce feelings of guilt if broken (Charness & Dufwenberg 2006).

(vi) There is even substantial cooperation in pure one-shot public goods games without any repetition (Marwell & Ames 1979; Gächter *et al.* 2004; Walker & Halloran 2004; Dufwenberg *et al.* 2006; Gächter & Herrmann 2007; Cubitt *et al.* 2008). This evidence is consistent with strong positive reciprocity. In §4, we will present an experimental design that sheds new light on strong positive reciprocity in the context of voluntary cooperation.

An important observation in all repeatedly played games reported in (i)–(iv) is that people make high contributions initially but over time contributions dwindle to low levels. The decay of cooperation has been replicated numerous times and has also been observed across a variety of participant pools (Herrmann *et al.* 2008). What explains this almost inevitable outcome? One possibility is learning the free-rider incentives. However, one problem with this explanation is that in experiments with a surprise restart contributions start high again, which is inconsistent with a pure learning hypothesis (Andreoni 1988; Croson 1996; Cookson 2000). People might also have some willingness to cooperate due to feelings of 'warm glow' (which might explain restart effects) but are otherwise confused decision makers who need time to learn what is the optimal contribution for them. Palfrey & Prisbrey (1997) test this idea and find some support for warm glow and reduced confusion over time. A further explanation, long argued by social psychologists (e.g. Kelley & Stahelski 1970), is that many people are conditional cooperators, who in principle are willing to cooperate if others do so as well, but get frustrated if others do not pull their weight. Therefore, the breakdown of cooperation is due to 'frustrated attempts at kindness' (Andreoni 1995; p. 900).

There is now mounting evidence from psychological and economic experiments for the importance of conditional cooperation both in the laboratory and

the field (Gächter 2007). In experiments that elicited participants' beliefs about how much they think others will contribute, contributions are indeed positively correlated with beliefs (Dufwenberg *et al.* 2006; Croson 2007; Fischbacher & Gächter 2008; Neugebauer *et al.* in press). A correlation does of course not establish causation and it is perfectly possible that a false consensus effect induces people to believe that others contribute the same as them (e.g. Kelley & Stahelski 1970). To circumvent this problem, Fischbacher *et al.* (2001) developed an experimental design in which the contribution of others was fixed. In their design, people have to indicate how much they contribute to the public good as a function of all possible average contribution levels of other group members. The results show that approximately 50 per cent are conditional cooperators, who increase their contributions if others contribute more, whereas approximately 25 per cent are free riders who never contribute anything—irrespective of how much others contribute. The rest show more complicated patterns.[8] Fischbacher & Gächter (2008) use the same method as Fischbacher *et al.* (2001) and show that the interaction of differently motivated people explains the decay of cooperation. The significance of this finding is that the decay of cooperation will occur not just because people eventually learn what is in their best interest but because frustrated conditional cooperators reduce their contributions. Thus, after some time, all types behave as income-maximizing free riders, even though only the free rider types are motivated by income maximization alone.

The fact that many people are conditional cooperators but some are free riders has two important general implications. First, the interaction structure matters (e.g. Gächter & Thöni 2005; Gunnthorsdottir *et al.* 2007), i.e. there is an 'ecology of collective action' (Ones & Putterman 2007). For instance, if cooperators know that they are among other 'like-minded' cooperators, they are able to maintain very high levels of cooperation (Gächter & Thöni 2005). Second, because conditional cooperators will adjust their cooperative behaviour to those observed around them and to what they believe others will do, any factor that shifts people's beliefs will shift their behaviour.[9]

Reciprocity is a likely source of conditional cooperation (Rabin 1993; Dufwenberg *et al.* 2006).[10] The reason is that cooperating is a nice act towards the other group members and people may want to return the favour. By contrast, free riding is an unkind act which people may want to punish. However, in the public goods experiments described above, the only way to punish free riding is to withdraw cooperation, with the consequence that other cooperators in the group get punished as well. This raises two questions: will people be willing to punish if they could target a free rider directly? Will the possibility to punish affect cooperation? Numerous experiments since the seminal studies of Yamagishi (1986) and Ostrom *et al.* (1992) have given affirmative answers to both the questions.

A typical design of most recent studies is as follows (Fehr & Gächter 2000, 2002). After participants have made their contribution decisions, group members are informed about how much the other group members have contributed to the public good. Each group member can then decide to punish each of the other group members. A punishment decision is implemented by assigning between 0 and 10 points to the punished member. Each point assigned reduces the punished member's income by $k \geq 1$ tokens and costs the punishing member one token. Punishment decisions are also made simultaneously and people are not informed about who punished them. Note that a rational and money-maximizing individual will never punish (in a one-shot game) because punishment is costly.

Numerous experiments have been conducted in this framework. Some of the results that are particularly interesting from the viewpoint of proximate and evolutionary theories of cooperation are as follows.

(i) Many people punish those who contribute less than them to the public good. In particular, the more someone free rides, the more he or she gets punished on average. This observation has been made in all public goods experiments with punishment we are aware of; there also seems to be little cross-cultural variation in the extent to which people punish freeloaders (Herrmann *et al.* 2008). Together with the cross-cultural evidence from ultimatum games and third party punishment games conducted in complex large-scale and small-scale societies around the globe (Oosterbeek *et al.* 2004; Henrich *et al.* 2005, 2006; Marlowe *et al.* 2008), these observations suggest that punishment of selfish behaviour is a 'human universal'.

(ii) The large majority of studies find that peer punishment increases and stabilizes cooperation at higher levels than without punishment. This is an important finding because the cooperation-enhancing effect of punishment is predicted by both proximate and ultimate theories of cooperation and punishment (Boyd *et al.* 2003; Fehr & Schmidt 2006; Carpenter *et al.* in press). There are exceptions, however. For instance, punishment does not work well if it is perceived as being unfair (e.g. van Prooijen *et al.* 2008) or if the group structure is asymmetric (Reuben & Riedl in press). There are also cross-cultural differences in the extent to which punishment establishes cooperation (see Herrmann *et al.* (2008) and §§4 and 5 of this paper).

(iii) The strategic nature of interaction (repeated interaction versus one-shot interaction) matters for cooperation but not much for punishment (Fehr & Gächter 2000). Put differently, while cooperation rates are significantly and substantially higher in repeated interactions when compared with repeated one-shot interactions, people punish free riding similarly, irrespective of whether it occurs in a repeated relationship or in random one-shot interactions. Moreover, as we will see in §4, people punish even in strict one-shot games with no repetition. Punishment is also often

harshest in the final period after people had experienced as many as 50 rounds of cooperation and punishment (Gächter *et al.* 2008). Any learning about the selfish incentives of the game should have taken place by then. Thus, these observations suggest that the level of cooperation is influenced by strategic considerations (free riding is less likely in repeated interactions), whereas punishment is to a large part non-strategic. Punishment seems to be an impulse triggered by negative emotions (Pillutla & Murnighan 1996; Bosman & van Winden 2002; Fehr & Gächter 2002; Sanfey *et al.* 2003; de Quervain *et al.* 2004; Knoch *et al.* 2006; Ben-Shakhar *et al.* 2007; Fehr & Camerer 2007; Seymour *et al.* 2007; Reuben & van Winden 2008) and not much by forward-looking considerations.

(iv) Although punishment is most likely to a large extent non-strategic and not forward looking, it follows economic rationality (cost–benefit considerations) in the sense that punishment is less likely used the more costly it is for the punishing individual (Anderson & Putterman 2006; Carpenter 2007*a*; Egas & Riedl 2008). The monitoring frequency and the severity of punishment inflicted on the punished individual also matters for the effectiveness of punishment to stabilize (or increase) cooperation (Carpenter 2007*b*; Egas & Riedl 2008; Nikiforakis & Normann 2008).

(v) There exists an interaction effect between the availability of punishment opportunities and direct reciprocity at the cooperation stage within stable groups. A repeated interaction and punishment are mutually reinforcing means to achieve high cooperation (e.g. Fehr & Gächter 2000; Masclet *et al.* 2003). If only direct reciprocity is possible, cooperation collapses, albeit it is higher than in random interactions. If only punishment is possible but groups are formed randomly and hence direct reciprocity is not feasible, cooperation is stabilized at intermediate levels. One reason why this is so is that punishment gives selfish individuals an incentive to cooperate and therefore also reinforces the beliefs of conditional cooperators that others will cooperate (Shinada & Yamagishi 2007). The experiment by Rockenbach & Milinski (2006) suggests that indirect reciprocity and punishment mutually reinforce cooperation as well. The advantage of direct and indirect reciprocity is that both help keeping the absolute costs of punishment low because they provide additional reasons to cooperate, and therefore reduce the need to maintain cooperation by costly punishment.[11]

(vi) Interestingly, punishment can also increase cooperation if it is purely symbolic and merely expresses social disapproval, without any material consequences for the punished individual (Masclet *et al.* 2003; Carpenter *et al.* 2004; Noussair & Tucker 2005). This suggests that punishment also triggers feelings of guilt and shame that induce individuals to behave pro-socially (Barr 2001; Fessler & Haley 2003). Hopfensitz & Reuben (in press) provide direct evidence for the role of shame and guilt in response to being punished. However, recent cross-cultural experiments suggest that punishment might not trigger guilt and shame in the same way everywhere, because in some participant pools punishment does not induce freeloaders to increase their contributions (Gintis 2008; Herrmann *et al.* 2008).

(vii) In most experiments in which punishment has material pay-off consequences, punishment turned out to be an inefficient tool to enforce cooperation because resources are destroyed. Indeed, in most experiments—which typically ran for 10 periods or less—net pay-offs in treatments with punishment were often lower than in treatments without punishment (e.g. Fehr & Gächter 2000; Page *et al.* 2005; Bochet *et al.* 2006; Botelho *et al.* 2007; Sefton *et al.* 2007; Dreber *et al.* 2008; Egas & Riedl 2008; Herrmann *et al.* 2008; Masclet & Villeval 2008; Nikiforakis 2008). For instance, Herrmann *et al.* (2008) has reported public goods experiments with and without punishment conducted in 16 comparable participant pools around the world. With the exception of three participant pools, the average pay-off in the experiments with punishment opportunities was lower than without punishment; and in those three participant pools with higher pay-offs, the increase was modest and amounted to 9.1, 2.8 and 0.5 per cent, respectively. Thus, 13 participant pools would have been better off not having had a punishment opportunity. The detrimental consequences of punishment are even more conspicuous if 'counter-punishment', i.e. multiple rounds of punishment, is possible (Denant-Boemont *et al.* 2007; Nikiforakis 2008).

(viii) The observation that punishment leaves groups worse off compared with experiments without punishment raises several interesting questions. For instance, Dreber *et al.* (2008) replicated the finding of the inefficiency of punishment in prisoner's dilemma experiments, and argue with reference to evolutionary (group-selection) models of altruistic punishment (in particular Boyd *et al.* 2003) that '[P]unishment therefore has no benefit for the group, which makes it hard to argue that punishment might have evolved by group selection' (p. 349). However, the observation that punishment is detrimental for group pay-offs stems predominantly from experiments that ran for 10 periods or less. Since punishment is to a large extent emotional and not forward looking, and because punishment is particularly used when cooperation is low, which typically is the case at the beginning of the experiment, the beneficial effects of punishment need more time to show up. Gächter *et al.* (2008) tested this possibility in experiments that ran for 50 periods and they compared pay-offs with those in 10-period experiments. As in previous

experiments, in the 10-period experiments punishment was detrimental in terms of payoffs when compared with 10-period experiments without punishment. In the 50-period experiments the opposite conclusion holds—cooperation is high and punishment costs negligible. Thus, if the time horizon is long enough, punishment can be group beneficial, a finding that supports models of group selection (Sober & Wilson 1998; Henrich & Boyd 2001; Boyd *et al.* 2003; Bowles 2006; Bowles & Choi 2007). A second interesting question is whether people would adopt a sanctioning institution if they had a choice. Gürerk *et al.* (2006) answered this question affirmatively, but there is an interesting twist. At the beginning of the experiments, people predominantly chose the non-sanctioning institution. As usual, there was substantial free riding, which tipped many people over to the punishment institution. Punishment then became the predominant choice for almost all people and very high levels of cooperation were established. Again, humans can also often communicate and coordinate punishment (Boehm 1993; Wiessner 2005; Reuben & van Winden 2008), which can minimize punishment costs. Finally, people can also frequently choose with whom to associate. Experiments show that both communication (Bochet *et al.* 2006) and voluntary association (Page *et al.* 2005) are indeed effective means to avoid the detrimental effects of punishment.

(ix) Given that people are willing to incur costs to punish others, would they also be willing to incur costs to reward others and would rewards (which are not efficiency reducing) steer people towards high contributions? Sefton *et al.* (2007) investigated this question in a design in which people could mutually reward each other such that a reward was a mere transfer of money from the rewarding subject to the rewarded subject. They compared this with punishment, i.e. a situation in which one punishment point assigned reduced the punished participant's income by one money unit whereas the punisher had to incur a cost of one. It turned out that people are prepared to reward cooperators, but punishment is more effective to increase contributions than rewards (see also Sutter *et al.* (2008) who got a similar result in a related design). The problem with rewards is that they need to be used when cooperation occurs, whereas punishment can work as a mere threat and need not be used much if people cooperate.

(x) Of particular relevance for evolutionary theories of cooperation are experiments where any future interaction with the same group members is excluded by design (so-called 'perfect stranger' matching). The reason why this is interesting is that the theories of direct and indirect reciprocity can explain why selfish people cooperate in repeated games with the same players but these theories predict little cooperation in one-shot games, because punishment is costly and bears no future benefits. To test this prediction, Fehr & Gächter (2002) conducted six rounds of anonymous public goods experiments with punishment under the perfect stranger matching design. In contrast to predictions, they observed substantial punishment of free riders in all rounds. Punishment under these circumstances is therefore evidence for strong negative reciprocity. Punishment is 'altruistic' because it is costly for the punisher, but due to the changed group composition in each round a punisher has no chance to benefit if the punished individual subsequently increases his or her contribution; only others benefit.[12] People punish others even in strict one-shot games without any repetition (Walker & Halloran 2004; Gächter & Herrmann 2007; Cubitt *et al.* 2008). In §4, we will provide further comprehensive evidence for strong negative reciprocity as it occurs in strict one-shot games.

In summary, there can be no doubt that direct and indirect reciprocity strongly shape human cooperation. However, there is also substantial cooperation when these channels are not available. We turn to this observation in §4.

4. STRONG RECIPROCITY AND CULTURAL BACKGROUND

In this section, we present an experiment that sheds new light on strong positive and negative reciprocity. This experiment also investigates how the cultural background influences patterns of both strong positive and negative reciprocity. The evidence on strong positive and negative reciprocity reviewed in §3 has contributed to the development of ultimate (e.g. Boyd *et al.* 2003) and proximate theories of why people cooperate and punish (see Fehr & Schmidt (2006) for a survey). Among the most important proximate psychological mechanisms are concerns for equity (Loewenstein *et al.* 1989; Dawes *et al.* 2007), and the punishment of kind and unkind intentions (Falk *et al.* 2005; Houser *et al.* 2008). These theories assume implicitly that motivations for strong reciprocity are similar across cultures (on average). Two reasons make it likely that the cultural environment exerts an influence on strong reciprocity, however. First, people have an innate ability to learn from others (Boyd & Richerson 1985; Tomasello *et al.* 2005). Cultural learning mechanisms will cause members of social groups to adopt similar values and beliefs about how others around them will reward and punish their behaviour (Sober & Wilson 1998; Henrich & Henrich 2007). Second, both strong positive and negative reciprocity might be shaped by local social norms about what constitutes the appropriate reaction to a benefit or harm one has received from others (Gouldner 1960; Coleman 1990; Sober & Wilson 1998; Henrich & Henrich 2007).

We are not the first to study cultural influences on strong reciprocity (seminal studies are by Henrich *et al.* 2005, 2006).[13] However, our methodology

differs in several important ways from previous approaches. First, we conducted public goods experiments with and without punishment, whereas previous studies mainly investigated bargaining games or third-party punishment games. Our set of games also allows us to study strong positive and negative reciprocity within one framework. In the context of our games, a strong reciprocator is predisposed to punish the non-cooperators (strong negative reciprocity) and to cooperate if others cooperate (strong positive reciprocity).

Second, we conducted our experiments one shot, anonymously and with people who did not know each other (the average participant had known only 6 per cent of other participants), because we wanted to measure strong reciprocity in a situation that was not confounded with reputational or strategic considerations coming from repeated play (Milinski *et al.* 2002; Fehr & Fischbacher 2003; Rockenbach & Milinski 2006).

Third, we elicited beliefs about how much others will contribute and how much they will punish. Owing to the one-shot nature of our experiments, participants deliberately could not base their expectation about how others were likely to behave on any observation made in the experiment. Participants had to form their expectations based on their experiences in daily life outside the laboratory. When we elicited beliefs we also asked participants how confident on a 10-point scale (1, very unconfident; 10, very confident) they were about their estimate. This is a measure of how precise people think their estimate is.

Fourth, we conducted our experiments in two highly developed industrialized countries (Russia and Switzerland). We are not interested in these countries *per se*, but they make interesting test cases as the 'cultural distance' between these societies is almost the largest one compared with all developed societies from which data are available.[14] We ran the Russian experiments in Belgorod and Yekaterinburg and the Swiss experiments in St Gallen and Zurich.[15] If the wider societal and cultural background influences patterns of strong reciprocity then it should affect beliefs and behaviour similarly in the two participant pools within a society and differently between societies.[16]

The specifics of our design are as follows. Groups of three participants played an anonymous one-shot public goods game (with $\alpha = 0.5$). We had two treatment conditions, one with no punishment opportunities (called the 'N-experiment', to measure strong positive reciprocity) and one with punishment opportunities ('the P-experiment', to measure strong negative reciprocity).

All participants took part in both a one-shot N-experiment and a one-shot P-experiment. We had two sequences: the N–P sequence, in which participants first played the N-experiment and then the P-experiment; in the P–N sequence this order was reversed. In both sequences, participants were unaware about the second experiment until they had finished the first one. This ensures the one-shot nature of the first experiments. We will therefore measure strong positive reciprocity in the N-experiment of the N–P sequence and strong negative reciprocity in the P-experiment of the P–N sequence. The reason for the two sequences is to see how participant pools react when punishment opportunities are added (in the N–P sequence), or removed (in the P–N sequence). Moreover, we can compare cooperation in the N-experiment of the N–P sequence with the P-experiment of the P–N sequence to see to what extent people anticipate the presence of a punishment option in their cooperation behaviour without any prior experience of the cooperativeness of others. A total of 603 people (360 Russian and 243 Swiss students) participated in either the N–P sequence ($n = 336$) or the P–N sequence ($n = 267$).

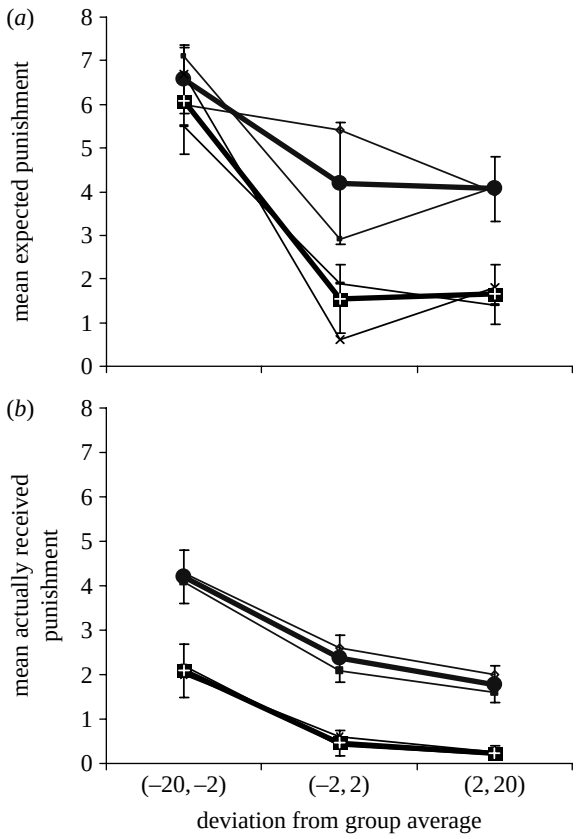

Figure 1. Cultural influences on strong negative reciprocity. (*a*) Mean punishment expected and (*b*) mean punishment received from other group members for a given deviation of own contribution from the group average. The error bars indicate the bootstrapped 95 per cent confidence bounds for country averages. Diamonds, Belgorod; squares, Yekaterinburg; circles, Russia; minus symbols, St Gallen; crosses, Zurich; plus symbols, Switzerland.

Figure 1*a* shows that in a case where a group member's contribution was lower than the group average contribution, expected punishment was very similar across participant pools (Kolmogorov–Smirnov test (KS test), $p = 0.821$). However, very strong differences between participant pools emerge in a case where a subject made similar contributions to those of his or her group members or contributed even more. In both cases we find that the Russian participant pools expected much more severe punishment than their Swiss counterparts. While the Swiss participants expected to receive 1.5 punishment points on average (with no significant differences (at $\alpha = 0.05$) between the two Swiss participant pools), their

Russian counterparts expected to receive almost 4.5 punishment points (also with no significant differences (at $\alpha = 0.05$) between participant pools). This difference is highly significant (KS test, $p < 0.004$).

Although participant pools held very different beliefs about the punishment they expected from their group members, people in all participant pools were similarly confident about their estimate. The average subject reports a confidence level of 6.03 and significantly more participants have a confidence level in the upper half than in the lower half of the scale (two-sided binomial test, $p = 0.005$).

Actual punishment (figure 1*b*) also shows a striking difference between the Swiss and the Russian participant pools. There are no significant differences (at $\alpha = 0.05$) between the participant pools within a society. However, punishment is highly significant and substantially harsher in the Russian than in the Swiss participant pools. This holds true for all deviation intervals (KS test, $p < 0.005$). The Russian participant pools punished not only the low contributors more severely than the Swiss participant pools, but also those who contributed at least as much as the group average. In the Swiss participant pools, punishment was almost exclusively directed at the low contributors. Thus, the cultural differences in actual punishment are not only in the severity with which people punish low contributors, but also in the way they punish high contributors. Such 'antisocial punishment' (Herrmann *et al.* 2008) is particularly puzzling, given that our one-shot design excludes retaliation (Herrmann *et al.* 2008; Nikiforakis 2008) for punishment received in the past as an explanation.

Are there also cultural influences on strong positive reciprocity as measured in the N-experiments of the N–P sequence?

As figure 2*a* shows, beliefs about others' contributions are not significantly different either between societies or between participant pools within societies (KS test, $p > 0.489$). Strong positive reciprocity in our one-shot game requires that people who believe that others make a high (low) contribution will reciprocate by contributing a high (low) amount as well (Fischbacher *et al.* 2001; Fehr & Fischbacher 2003; Dufwenberg *et al.* 2006; Croson 2007). Thus, beliefs about others' contributions and own contributions should be positively correlated. This is indeed the case in all participant pools (figure 2*b*). However, despite the fact that beliefs are not significantly different between participant pools, we also find cultural influences on strong positive reciprocity in the sense that the relationship between contributions and beliefs is steeper in both the Swiss pools than in both the Russian pools. The main reason for this difference is that contributions towards high beliefs about others (expected contributions in the interval (14, 20)) are substantially lower in the Russian participant pools than in the Swiss participant pools (KS test, $p = 0.001$); no significant differences can be detected in the other intervals (KS test; $p > 0.113$).

The cultural differences in strong reciprocity also had an impact on cooperation (figure 3). In the N-experiment of the N–P sequence, the resulting contributions levels were significantly lower in the

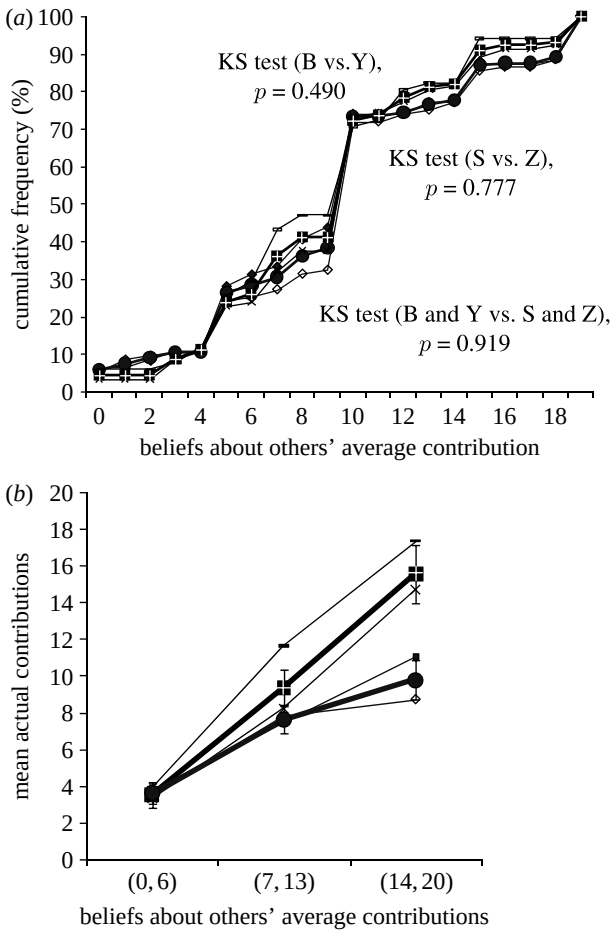

Figure 2. Cultural influences on strong positive reciprocity. (*a*) Distribution of beliefs about the average contribution of the other two group members, separately for each participant pool and pooled for the Russian and the Swiss participant pools, respectively. KS test indicates Kolmogorov–Smirnov tests about the equality of distributions. (*b*) Mean actual contribution of a given belief about others' contribution. Error bars indicate bootstrapped 95 per cent confidence bounds of country averages. Diamonds, Belgorod (B); squares, Yekaterinburg (Y); circles, Russia; minus symbols, St Gallen (S); crosses, Zurich (Z); plus symbols, Switzerland.

Russian than the Swiss participant pools (KS test, $p < 0.006$); there were no significant differences within societies (KS test, $p > 0.143$). Similarly, in the P-experiment of the P–N sequence, contributions of the Swiss participant pools were significantly higher than the Russian participant pools (KS test, $p < 0.001$). As in the N-experiment, there are virtually no differences in the distribution of contributions within both the Russian and the Swiss participant pools (KS test, $p > 0.659$).

As a consequence of different cooperation and punishment patterns, earnings in the P-experiment are highly significantly different between the Russian and the Swiss participant pools, but not significantly different within societies. Eighty per cent of the Russian participants earned less than 20 money units—the earnings predicted for selfishly rational players. In Switzerland, this was true for 33 per cent of participants.

Our final steps are, first, to compare contributions to the N-experiments of the N–P sequence and the P-experiments of the P–N sequence. This analysis informs us about the extent to which participants

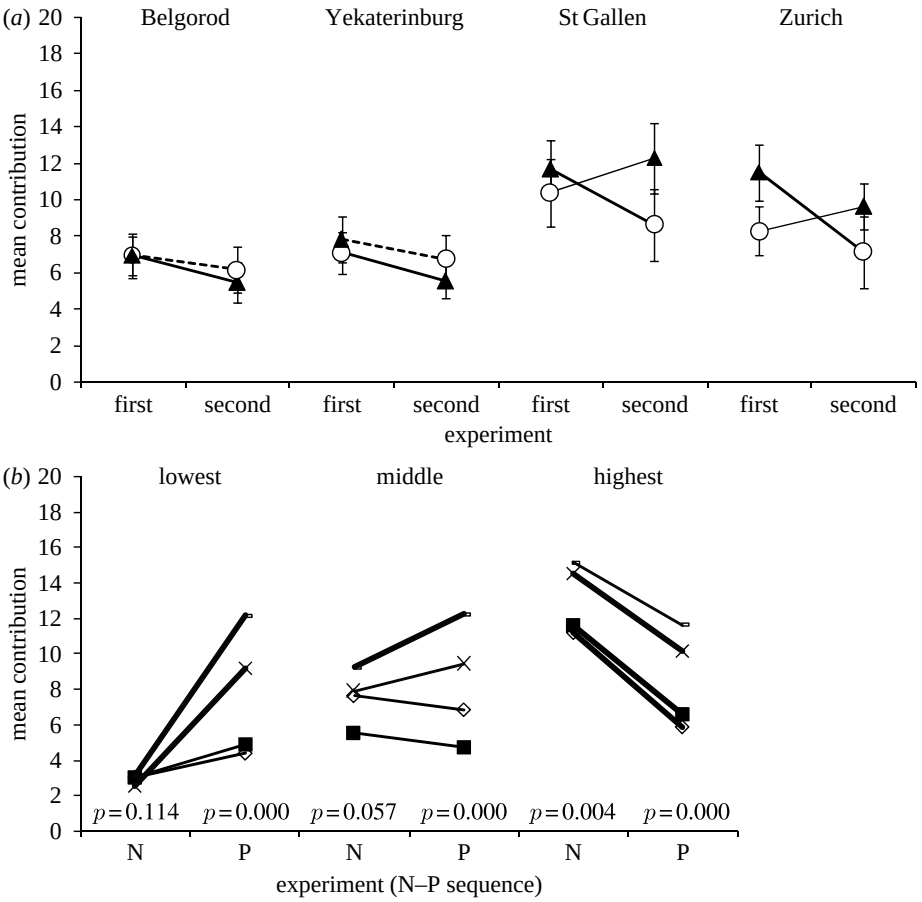

Figure 3. Cultural differences on the impact of strong reciprocity on cooperation. (*a*) Change in contributions when punishment is added (in the N–P sequence) or subtracted (in the P–N sequence). The thickness of the connecting lines indicates the significance level of the behavioural change according to Wilcoxon matched-pair tests (with group average contributions as independent observations): ------, $p > 0.1$; ——, $p < 0.05$; ——, $p < 0.01$. The error bars are the bootstrapped 95 per cent confidence bounds of mean contributions (circles, N-experiment; triangles, P-experiment). (*b*) Change in contribution in the P-experiment compared with the N-experiment in the N–P sequence by the groups' minimum, middle and maximum contributors of the N-experiment. We indicate the *p*-values of Kruskal–Wallis tests of equality of contributions in all four participant pools. Diamonds, Belgorod; squares, Yekaterinburg; minus symbols, St Gallen; crosses, Zurich.

anticipate punishment in their contribution behaviour without any prior experience of others' behaviour. Second, we look at the change in contributions to the N–P sequence, where we introduce a punishment opportunity after participants have had some experience with cooperation behaviour in the N-experiment.

Zurich is the only participant pool where contributions are significantly higher in the P-experiment than in the N-experiment (KS test, $p = 0.006$; comparing the first experiments in a sequence). In the other participant pools, contributions are only insignificantly higher (Yekaterinburg and St Gallen; KS test, $p > 0.215$) or even slightly lower (Belgorod, KS test, $p = 0.996$).

In the P–N sequence, contributions from all four participant pools are highly significantly lower in the N-experiment than in the preceding P-experiment. By contrast, in the N–P sequence in both the Swiss participant pools, contributions to the P-experiment are significantly higher than the N-experiment. The opposite is true in both the Russian participant pools.[17]

To shed light on the cultural differences in the dynamics of cooperation when a punishment option is added, we look at individual group members in the N-experiments of the N–P sequence and investigate how they change their contribution to the P-experiment. We classify each group member in the N-experiment whether he or she is the lowest, middle or highest contributor in his or her group (figure 3*b*). The lowest contributors in the N-experiment in the Swiss participant pools increased their contributions to the P-experiment substantially (by 6.83 tokens on average), whereas in Russia the lowest contributors raised their contribution to the P-experiment only modestly (by 1.60 tokens on average). Similarly, the middle contributors raised their contributions to both the Swiss participant pools, whereas in the Russian participant pools contributions dropped. Surprisingly, the top contributors lowered their contributions in all four participant pools.

In summary, the experiment presented here unambiguously shows two things: first, people on average are strong reciprocators who cooperate if they believe others cooperate and punish free riders. Second, strong reciprocity, especially strong negative reciprocity, is subject to substantial cultural influences. A particularly noteworthy phenomenon is the anti-social punishment observed in the Russian participant pools—people punished not only the free riders but the cooperators too, and the latter even expected being

punished.[18] In the remaining two sections, we present evidence on how general the findings are along two important dimensions: different societies (§5) and different socio-economic groups (§6).

5. ANTISOCIAL PUNISHMENT ACROSS SOCIETIES

The results from §4 suggest that the cultural background matters for cooperation and punishment behaviour. Stimulated by this result, Herrmann *et al.* (2008) undertook a large-scale experiment across 16 different participant pools in 15 different societies around the world. In their experiments, groups of four played 10 periods of a public goods game without punishment followed by 10 periods without punishment. The results showed striking similarities as well as differences in punishment behaviour. The striking similarities occurred in the punishment of free-riding behaviour: across all subject pools people punished freeloaders very similarly. Large differences arose in the punishment of cooperators (antisocial punishment). In some subject pools antisocial punishment was virtually absent, whereas in others it was as prevalent as punishment of freeloaders. As a consequence, cooperation levels were vastly different: some participant pools invested almost all their endowment to the public good, whereas in others people invested less than a third. Punishment stabilized cooperation everywhere. In the experiment without punishment cooperation collapsed, as in almost all previous experiments.

What explains antisocial punishment? Tentative answers can be given at two levels. At a macro level, Herrmann *et al.* (2008) found that antisocial punishment occurred predominantly in societies with weak social norms of cooperation, weak rules of law and weak democracies, according to measures developed by various social scientists using representative survey data. At the individual level antisocial punishment may be motivated by revenge (Denant-Boemont *et al.* 2007; Nikiforakis 2008), at least in some societies (Herrmann *et al.* 2008; Mohan 2008). There might also be cultural differences in the extent to which people are motivated by relative pay-offs (Liebrand *et al.* 1986; Zizzo 2003; Fliessbach *et al.* 2007) and concerns for dominance (Clutton-Brock & Parker 1995). People might also dislike 'do-gooders' (Monin 2007), punish non-conformists (Carpenter & Matthews 2005) and punish displays of conspicuous generosity (Henrich *et al.* 2006). Some punishment might also be motivated by selfish considerations to induce others to contribute even more (Eldakar *et al.* 2007). Finally, punishment might be linked to the perception of group boundaries: some (traditional) societies are structured along strong private networks with a lot of cooperation within networks and little beyond. Because participants did not know each other (and were outside each others' networks), they might not have accepted punishment from an outsider. Punishment might trigger anger, not guilt (Gintis 2008). Indeed, antisocial punishment occurred predominantly in more traditional, segmentary societies. Which of these explanations is important is a task for future research.

6. SOCIO-DEMOGRAPHIC INFLUENCES ON STRONG RECIPROCITY

In most experiments discussed above, researchers had used participants who were similar in age, educational and socio-economic background; in the cross-cultural experiments, the rationale was to maximize comparability across participant pools. However, there is evidence that some socio-demographic characteristics (in particular, age) matter for social preferences (e.g. Fehr *et al.* 2002*b*; Carpenter *et al.* 2005*b*, 2008; Holm & Nystedt 2005; Bellemare & Kröger 2007; Sutter 2007; Sutter & Kocher 2007; Bellemare *et al.* 2008; Dohmen *et al.* 2008; Egas & Riedl 2008). This raises the question of whether the patterns of punishment observed above also hold for a more representative sample of the population, not just young people.

To test for the generalizability of our findings, we ran experiments very similar to those reported in §4 with 566 Russian urban and rural dwellers of all age cohorts (Gächter & Herrmann 2007). We were also interested in running the experiments in urban and rural areas, because the gap between them is particularly pronounced in Russia. Moreover, norm enforcement may be easier in close-knit rural communities than in anonymous urban areas (Bowles & Gintis 2002). We ran our experiments in the urban area of Kursk, a city in the heartland of the former Soviet Union, and in the rural areas surrounding Kursk. We had four participant pools: two mature pools ('urban mature' and 'rural mature'), i.e. people who on average were 44 years old and had spent most of their life in a big city (a rural area); and two young participant pools with an average age of 21 years ('urban young' and 'rural young'). The design was the same as the one described above. The only exception was that for practical purposes the experiments were hand run and we did not elicit beliefs.

The results strongly resemble the ones reported above. We found in all four participant pools not only high levels of punishment of people who contributed less than the punishing subject but also substantial antisocial punishment of people who contributed the same or even more. In no participant pool did punishment lead to an increase in cooperation. In particular, contributions in all four pools dropped even in the N–P sequence, as in the experiments reported above. None of the socio-demographic background variables matters for punishment but some of them matter for cooperation behaviour. In particular, rural dwellers were more cooperative than their urban counterparts and the older people were the more they contributed to the public good both in the N- and the P-experiment.

Our observation that age only matters for cooperation behaviour but not for punishment stands in contrast to findings from public goods experiments with and without punishment conducted with more than 800 Dutch people from all age cohorts (average age 35 years; Egas & Riedl 2008). They found that age was only (weakly) significantly (and not very robustly against other specifications) correlated with contributions. However, unlike in our Russian experiments, age was a significant predictor for punishment behaviour—the older people were the more they

punished others, *ceteris paribus*. Thus, the relevance of socio-demographic background variables may also be subject to cultural influences.

7. CONCLUDING REMARKS

From the experimental evidence we reviewed here, there can be no doubt that direct reciprocity (aka 'reciprocal altruism') and indirect reciprocity (helping those in good standing) are very important determinants of human cooperative behaviour. Yet, there is substantial accumulated evidence that people also cooperate and punish in anonymous one-shot games where future gains from cooperation, or reputational benefits, are excluded by design. We view the numerous observations of substantial co-operation and punishment in one-shot games as supporting evidence for strong reciprocity.

We believe that understanding strong reciprocity is of importance for a variety of behavioural disciplines for which cooperation (and culture) are central issues (Ostrom 1998; Fehr & Fischbacher 2003; Hagen & Hammerstein 2006; Sigmund 2007). The findings reviewed here, in particular those from the cross-cultural experiments, support anthropological and evolutionary theories of cooperation which predict that people's social preferences are programmable and therefore culturally variable (Henrich 2004; Henrich *et al.* 2005). Our results also demonstrate that to explain our patterns of strong reciprocity models of decision making in game theory, economics and psychology need to develop models of social preferences (e.g. concerns for equity and the reward and punishment of kind and unkind intentions (Falk *et al.* 2005)) that take cultural influences on those motivations into account. In particular, the role of cultural influences on strong negative reciprocity deserves extensive scrutiny as here the cultural differences appear to be largest (Herrmann *et al.* 2008). Previous explanations have focused predominantly on altruistic punishment of low contributors (Sigmund 2007). Our results show that there is also a need to understand why people punish those who behave pro-socially and what the cultural determinants of antisocial punishment are.

We are grateful to the Universities of Belgorod, Yekaterinburg, Zurich and St Gallen for their support in running the experiments. We also gratefully acknowledge financial support from the University of Nottingham, the Latsis Foundation (Geneva) and the EU-TMR Research Network ENDEAR (FMRX-CT98-0238). We received helpful comments from the referees and various workshop audiences, in particular the Arts and Humanities Research Council workshops Culture and the Mind in Sheffield, and from Jo Morgan and Daniel Scruton. S.G. also gratefully acknowledges the hospitality of the Center for Economic Studies in Munich and the Economics Department of the University of Sydney while working on this paper. This paper is part of the MacArthur Foundation Network on Economic Environments and the Evolution of Individual Preferences and Social Norms.

ENDNOTES

[1]In all the laboratory experiments we discuss, participants, depending on their decisions, earned considerable amounts of money. Thus, the laboratory allows observing real decision making under controlled circumstances. See Friedman & Sunder (1994) for an introduction to methods in experimental economics; Guala (2005) for a discussion of the methodology of experimental economics; and Kagel & Roth (1995) and Camerer (2003) for an overview of the important experimental results across a variety of human decision making problems.

[2]Conducting experiments in environments outside the university laboratory ('field experiments') is a burgeoning area in experimental economics. See e.g. Carpenter *et al.* (2005*a*) for an overview and some applications.

[3]Other important areas of empirical research in collective decision making concern coordination problems and problems of collective choice. For lack of space we do not discuss this research here. We refer the reader to Camerer (2003) and Devetag & Ortmann (2007) for recent surveys on coordination games, and Palfrey (2008) on experiments in collective choice. For further aspects of human collective decision making, see Austen-Smith & Feddersen (2009), Conradt & Roper (2009), Dyer *et al.* (2009), Hix (2009) and Skyrms (2009).

[4]The prisoner's dilemma is another useful tool for studying cooperation. It was particularly popular in early experimental research on cooperation. See Rapoport & Chammah (1965) and Colman (1999) for overviews, and Dreber *et al.* (2008) for a recent example. The disadvantage of the prisoner's dilemma is that it is restricted to bilateral interactions, which have different theoretical properties from multi-lateral interactions, in particular in repeated interactions (e.g. Boyd & Richerson 1988).

[5]For overviews, see Dawes (1980), Ledyard (1995), Kollock (1998), Zelmer (2003), Gächter & Herrmann (2005) and Gächter (2007).

[6]There are some exceptions. See e.g. Andreoni (1988), Weimann (1994) and Andreoni & Croson (2008) for an overview.

[7]Cues of kinship also increase cooperation (Madsen *et al.* 2007; Krupp *et al.* 2008).

[8]Herrmann & Thöni (in press) and Kocher *et al.* (2008) replicated the Fischbacher *et al.* (2001) study using the same parameters. They got similar results. See Kurzban & Houser (2005), Bardsley & Moffatt (2007) and Muller *et al.* (2008) for related studies that also report substantial individual differences in cooperative attitudes. See Doebeli *et al.* (2004) for an evolutionary explanation of type heterogeneity.

[9]See Gächter (2007) for several examples and a general discussion.

[10]Conformity is another source of conditional cooperation—people just do what others do. Carpenter (2004) and Bardsley & Sausgruber (2005) provide evidence for the relevance of conformity in voluntary cooperation. See Gächter (2007) for an overview of studies on conditional cooperation and discussions of related issues.

[11]Another mechanism to keep the costs of altruistic punishment low is when punishment leads to a reputational benefit for the punisher. See Barclay (2006) for a study that suggests this possibility.

[12]Egas & Riedl (2008) replicated this result with a large number of Dutch residents across all age cohorts and various socio-demographic backgrounds.

[13]See the supplementary materials for further references to cross-cultural experiments.

[14]We conducted our experiments in Yekaterinburg and Belgorod (Russia) and St Gallen and Zurich (Switzerland). Both countries are highly industrialized, rely on large-scale division of labour and have extensive trade among genetically unrelated strangers. Compared to the small-scale societies of previous studies (Henrich *et al.* 2005, 2006), the main distinguishing features between Russia and Switzerland are therefore not in the fundamentals of socio-economic organization but in historical, religious, political and cultural values, which are hugely different between these societies according to frequently used measures developed by various social scientists interested in quantifying cultural and societal differences (Inglehart & Baker 2000; Hofstede 2001). The cultural distance between Switzerland and Russia (measured as the Euclidean distance between country scores of the respective indicators) is almost the largest one compared with the 55 countries from which data are available. See the electronic supplementary material for further details.

[15]Belgorod is a medium-sized city (roughly 300 K inhabitants) in the southwest of Russia, near the border to Ukraine. Yekaterinburg is a big city (more than 1000 K inhabitants) in the Ural region, 1000 miles east of Moscow. These cities are representative of Russia outside Moscow. Zurich is located in the centre of Switzerland and

its urban area has roughly 1000 K inhabitants. St Gallen has roughly 80 K inhabitants and is the major centre in the Northeast region of Switzerland. Both cities are representative of German-speaking Switzerland.

[16]To maximize comparability across participant pools, we implemented the following procedures: (i) We had all instructions translated into Russian, and back-translated, to control for language-induced differences in meaning; (ii) All instructions were written in a neutral language, to avoid evoking culture-specific meanings; (iii) We followed exactly the same protocol in the manner in which we conducted the experiments in all participant pools—in particular, participants had to answer the same set of control questions that tested their understanding of payoff calculations before the experiment could start; (iv) We conducted all experiments with people who did not know each other and (v) in computerized laboratories in which participants were visually separated from one another to ensure between-subject anonymity and to maximize subject–experimenter anonymity; (vi) We used the same software (Fischbacher 2007), i.e. participants saw the same interface (except for different languages); (vii) During the experiment, we calculated all incomes in 'Guilders', to avoid number and currency effects whose perception might differ across cultures; (viii) We used the same stake size in relative monthly income, i.e. we chose the exchange rate between 'Guilders' and the local currency such that real expected earnings were roughly the same; and (viii) to minimize experimenter effects, the same experimenter (B. Herrmann, who speaks German and Russian fluently) organized and supervised all 25 sessions according to exactly the same script. See the supplementary information (available upon request) for further details.

[17]With regard to strong positive and negative reciprocity (and its impact on cooperation and earnings), we get very similar results in the second experiments of our N–P and P–N sequence. Thus, our findings are robust to order effects.

[18]Cinyabuguma *et al.* (2006) call the punishment of cooperators 'perverse punishment'.

REFERENCES

Anderson, C. M. & Putterman, L. 2006 Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games Econ. Behav.* **54**, 1–24. (doi:10.1016/j.geb.2004.08.007)

Anderson, S. P., Goeree, J. K. & Holt, C. A. 1998 A theoretical analysis of altruism and decision error in public goods games. *J. Public Econ.* **70**, 297–323. (doi:10.1016/S0047-2727(98)00035-8)

Andreoni, J. 1988 Why free ride—strategies and learning in public-goods experiments. *J. Public Econ.* **37**, 291–304. (doi:10.1016/0047-2727(88)90043-6)

Andreoni, J. 1995 Cooperation in public-goods experiments—kindness or confusion? *Am. Econ. Rev.* **85**, 891–904.

Andreoni, J. & Croson, R. 2008 Partners versus strangers: the effect of random rematching in public goods experiments. In *The handbook of experimental economic results* (eds C. Plott & V. Smith), pp. 776–783. Amsterdam, The Netherlands: Elsevier.

Andreoni, J. & Petrie, R. 2004 Public goods experiments without confidentiality: a glimpse into fund-raising. *J. Public Econ.* **88**, 1605–1623. (doi:10.1016/S0047-2727(03)00040-9)

Austen-Smith, D. & Feddersen, T. J. 2009 Information aggregation and communication in committees. *Phil. Trans. R. Soc. B* **364**, 763–769. (doi:10.1098/rstb.2008.0256)

Axelrod, R. & Hamilton, W. 1981 The evolution of cooperation. *Science* **211**, 1390–1396. (doi:10.1126/science.7466396)

Barclay, P. 2006 Reputational benefits for altruistic punishment. *Evol. Hum. Behav.* **27**, 325–344. (doi:10.1016/j.evolhumbehav.2006.01.003)

Bardsley, N. & Moffatt, P. G. 2007 The experimetrics of public goods: inferring motivations from contributions. *Theory Decision* **62**, 161–193. (doi:10.1007/s11238-006-9013-3)

Bardsley, N. & Sausgruber, R. 2005 Conformity and reciprocity in public good provision. *J. Econ. Psychol.* **26**, 664–681. (doi:10.1016/j.joep.2005.02.001)

Barr, A. 2001 Social dilemmas and shame-based sanctions: experimental results from rural Zimbabwe. Center for the Study of African Economies working paper. WPS2001.11.

Bateson, M., Nettle, D. & Roberts, G. 2006 Cues of being watched enhance cooperation in a real-world setting. *Biol. Lett.* **2**, 412–414. (doi:10.1098/rsbl.2006.0509)

Bellemare, C. & Kröger, S. 2007 On representative social capital. *Eur. Econ. Rev.* **51**, 183–202. (doi:10.1016/j.euroecorev.2006.03.006)

Bellemare, C., Kröger, S. & Van Soest, A. 2008 Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica* **76**, 815–839. (doi:10.1111/j.1468-0262.2008.00860.x)

Ben-Shakhar, G., Bornstein, G., Hopfensitz, A. & van Winden, F. 2007 Reciprocity and emotions in bargaining using physiological and self-report measures. *J. Econ. Psychol.* **28**, 314–323. (doi:10.1016/j.joep.2007.02.005)

Bochet, O., Page, T. & Putterman, L. 2006 Communication and punishment in voluntary contribution experiments. *J. Econ. Behav. Organ.* **60**, 11–26.

Boehm, C. 1993 Egalitarian behavior and reverse dominance hierarchy. *Curr. Anthropol.* **34**, 227–254. (doi:10.1086/204166)

Bosman, R. & van Winden, F. 2002 Emotional hazard in a power-to-take experiment. *Econ. J.* **112**, 147–169. (doi:10.1111/1468-0297.0j677)

Botelho, A., Harrison, G. W., Costa Pinto, L. M. & Rutström, E. E. 2007 Social norms and social choice. Working paper no. 05-23, Economics Department, University of Central Florida.

Bowles, S. 2006 Group competition, reproductive leveling, and the evolution of human altruism. *Science* **314**, 1569–1572. (doi:10.1126/science.1134829)

Bowles, S. & Choi, J.-K. 2007 The coevolution of parochial altruism and war. *Science* **318**, 636–640. (doi:10.1126/science.1149498)

Bowles, S. & Gintis, H. 2002 Social capital and community governance. *Econ. J.* **112**, 419–436. (doi:10.1111/1468-0297.00077)

Boyd, R. & Richerson, P. J. 1985 *Culture and the evolutionary process*. Chicago, IL: University of Chicago Press.

Boyd, R. & Richerson, P. J. 1988 The evolution of reciprocity in sizable groups. *J. Theor. Biol.* **132**, 337–356. (doi:10.1016/S0022-5193(88)80219-4)

Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. 2003 The evolution of altruistic punishment. *Proc. Natl Acad. Sci. USA* **100**, 3531–3535. (doi:10.1073/pnas.0630443100)

Brandts, J. & Schram, A. 2001 Cooperation and noise in public goods experiments: applying the contribution function approach. *J. Public Econ.* **79**, 399–427. (doi:10.1016/S0047-2727(99)00120-6)

Brosig, J., Ockenfels, A. & Weimann, J. 2003 The effect of communication media on cooperation. *Ger. Econ. Rev.* **4**, 217–241. (doi:10.1111/1468-0475.00080)

Burnham, T. C. & Hare, B. 2007 Engineering human cooperation—does involuntary neural activation increase public goods contributions? *Hum. Nat. Interdiscip. Biosoc. Perspect.* **18**, 88–108.

Camerer, C. F. 2003 *Behavioral game theory*. Princeton, NJ: Princeton University Press.

Camerer, C. F. & Fehr, E. 2004 Measuring social norms using experimental games: a guide for social scientists. In *Foundations of human sociality: experimental and ethnographic evidence from 15 small-scale societies* (eds J. Henrich, R. Boyd, S. Bowles, C. F. Camerer, E. Fehr & H. Gintis), pp. 55–95. New York, NY: Oxford University Press.

Cardenas, J. C. & Jaramillo, C. R. 2007 *Cooperation in large networks. An experimental approach. Documento CEDE 2007-06* (*Edición Electrónica*). Bogotá, Colombia: Universidad de los Andes.

Carpenter, J. P. 2004 When in Rome: conformity and the provision of public goods. *J. Socio-Econ.* **33**, 395–408. (doi:10.1016/j.socec.2004.04.009)

Carpenter, J. 2007*a* The demand for punishment. *J. Econ. Behav. Organ.* **62**, 522–542. (doi:10.1016/j.jebo.2005.05.004)

Carpenter, J. 2007*b* Punishing free-riders: how group size affects mutual monitoring and the provision of public goods. *Games Econ. Behav.* **60**, 31–51. (doi:10.1016/j.geb.2006.08.011)

Carpenter, J. & Matthews, P. 2005 Norm enforcement: anger, indignation or reciprocity? IZA discussion paper no. 1583.

Carpenter, J. P., Daniere, A. G. & Takahashi, L. M. 2004 Cooperation, trust, and social capital in Southeast Asian urban slums. *J. Econ. Behav. Organ.* **55**, 533–551. (doi:10.1016/j.jebo.2003.11.007)

Carpenter, J., Harrison, G. & List, J. (eds) 2005*a* *Field experiments in economics*. Greenwich, CT: JAI Press.

Carpenter, J. P., Burks, S. & Verhoogen, E. 2005*b* Comparing students to workers: the effect of stakes, social framing, and demographics on bargaining outcomes. In *Field experiments in economics* (eds J. P. Carpenter, G. Harrison & J. List). Research in experimental economics. Amsterdam, The Netherlands: Elsevier JAI.

Carpenter, J., Connolly, C. & Knowles Myers, C. 2008 Altruistic behavior in a representative dictator experiment. *Exp. Econ.* **11**, 282–298. (doi:10.1007/s10683-007-9193-x)

Carpenter, J., Bowles, S., Gintis, H. & Hwang, S.-H. In press. Strong reciprocity and team production: theory and evidence. *J. Econ. Behav. Organ.*

Charness, G. & Dufwenberg, M. 2006 Promises and partnership. *Econometrica* **74**, 1579–1601. (doi:10.1111/j.1468-0262.2006.00719.x)

Chaudhuri, A., Graziano, S. & Maitra, P. 2006 Social learning and norms in a public goods experiment with inter-generational advice. *Rev. Econ. Stud.* **73**, 357–380. (doi:10.1111/j.1467-937X.2006.0379.x)

Cinyabuguma, M., Page, T. & Putterman, L. 2006 Can second-order punishment deter perverse punishment? *Exp. Econ.* **9**, 265–279. (doi:10.1007/s10683-006-9127-z)

Clutton-Brock, T. H. & Parker, G. A. 1995 Punishment in animal societies. *Nature* **373**, 209–216. (doi:10.1038/373209a0)

Cochard, F., Van, P. N. & Willinger, M. 2004 Trusting behavior in a repeated investment game. *J. Econ. Behav. Organ.* **55**, 31–44. (doi:10.1016/j.jebo.2003.07.004)

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; New York, NY: Routledge.

Conradt, L. & Roper, T. J. 2009 Conflicts of interest and the evolution of decision sharing. *Phil. Trans. R. Soc. B* **364**, 807–819. (doi:10.1098/rstb.2008.0257)

Cookson, R. 2000 Framing effects in public goods experiments. *Exp. Econ.* **3**, 55–79. (doi:10.1007/BF01669207)

Croson, R. 1996 Partners and strangers revisited. *Econ. Lett.* **53**, 25–32. (doi:10.1016/S0165-1765(97)82136-2)

Croson, R. 2007 Theories of commitment, altruism and reciprocity: evidence from linear public goods games. *Econ. Inq.* **45**, 199–216.

Cubitt, R., Drouvelis, M. & Gächter, S. 2008 Framing and free riding: emotional responses and punishment in social dilemma games. CeDEx discussion paper no. 2008-02, University of Nottingham.

Dal Bo, P. 2005 Cooperation under the shadow of the future: experimental evidence from infinitely repeated games. *Am. Econ. Rev.* **95**, 1591–1604. (doi:10.1257/000282805775014434)

Dawes, R. M. 1980 Social dilemmas. *Annu. Rev. Psychol.* **31**, 169–193. (doi:10.1146/annurev.ps.31.020180.001125)

Dawes, R. M., McTavish, J. & Shaklee, H. 1977 Behavior, communication, and assumptions about other peoples behavior in a commons dilemma situation. *J. Pers. Soc. Psychol.* **35**, 1–11. (doi:10.1037/0022-3514.35.1.1)

Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R. & Smirnov, O. 2007 Egalitarian motives in humans. *Nature* **446**, 794–796. (doi:10.1038/nature05651)

de Quervain, D. J. F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A. & Fehr, E. 2004 The neural basis of altruistic punishment. *Science* **305**, 1254–1258. (doi:10.1126/science.1100735)

Denant-Boemont, L., Masclet, D. & Noussair, C. N. 2007 Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Econ. Theor.* **33**, 145–167. (doi:10.1007/s00199-007-0212-0)

Devetag, G. & Ortmann, A. 2007 When and why? A critical survey on coordination failure in the laboratory. *Exp. Econ.* **10**, 331–344. (doi:10.1007/s10683-007-9178-9)

Doebeli, M., Hauert, C. & Killingback, T. 2004 The evolutionary origin of cooperators and defectors. *Science* **306**, 859–862. (doi:10.1126/science.1101456)

Dohmen, T., Falk, A., Huffman, D. & Sunde, U. 2008 Representative trust and reciprocity: prevalence and determinants. *Econ. Inq.* **46**, 84–90.

Dreber, A., Rand, D. G., Fudenberg, D. & Nowak, M. A. 2008 Winners don't punish. *Nature* **452**, 348–351. (doi:10.1038/nature06723)

Dufwenberg, M., Gächter, S. & Hennig-Schmidt, H. 2006 The framing of games and the psychology of strategic choice. CeDEx discussion paper 2006-20, University of Nottingham.

Dyer, J. R. G., Johansson, A., Helbing, D., Couzin, I. D. & Krause, J. 2009 Leadership, consensus decision making and collective behaviour in humans. *Phil. Trans. R. Soc. B* **364**, 781–789. (doi:10.1098/rstb.2008.0233)

Egas, M. & Riedl, A. 2008 The economics of altruistic punishment and the maintenance of cooperation. *Proc. R. Soc. B* **275**, 871–878. (doi:10.1098/rspb.2007.1558)

Eldakar, O. T., Farrell, D. L. & Wilson, D. S. 2007 Selfish punishment: altruism can be maintained by competition among cheaters. *J. Theor. Biol.* **249**, 198–205. (doi:10.1016/j.jtbi.2007.07.024)

Engelmann, D. & Fischbacher, U. 2002 Indirect reciprocity and strategic reputation building in an experimental helping game. IEW Working paper no. 132, University of Zurich.

Falk, A., Gächter, S. & Kovacs, J. 1999 Intrinsic motivation and extrinsic incentives in a repeated game with incomplete contracts. *J. Econ. Psychol.* **20**, 251–284. (doi:10.1016/S0167-4870(99)00009-4)

Falk, A., Fehr, E. & Fischbacher, U. 2005 Driving forces behind informal sanctions. *Econometrica* **73**, 2017–2030. (doi:10.1111/j.1468-0262.2005.00644.x)

Fehr, E. & Camerer, C. 2007 Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* **11**, 419–427. (doi:10.1016/j.tics.2007.09.002)

Fehr, E. & Fischbacher, U. 2003 The nature of human altruism. *Nature* **425**, 785–791. (doi:10.1038/nature02043)

Fehr, E. & Gächter, S. 2000 Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* **90**, 980–994.

Fehr, E. & Gächter, S. 2002 Altruistic punishment in humans. *Nature* **415**, 137–140. (doi:10.1038/415137a)

Fehr, E. & Schmidt, K. M. 2006 The economics of fairness, reciprocity and altruism—experimental evidence and new theories. In *Handbook of the economics of giving, altruism and reciprocity*, vol. 1 (eds S.-C. Kolm & J. M. Ythier), pp. 615–691. Amsterdam, The Netherlands: Elsevier B.V.

Fehr, E., Fischbacher, U. & Gächter, S. 2002*a* Strong reciprocity, human cooperation, and the enforcement of social norms. *Hum. Nat. Interdiscip. Biosoc. Perspect.* **13**, 1–25. (doi:10.1007/S12110-002-1012-7)

Fehr, E., Fischbacher, U., von Rosenbladt, B., Schupp, J. & Wagner, G. G. 2002*b* A nationwide laboratory. Examining trust and trustworthiness by integrating behavioral experiments into representative surveys. *Schmoller's Jahrbuch* **122**, 519–542.

Fessler, D. & Haley, K. J. 2003 The strategy of affect: emotions in human cooperation. In *Genetic and cultural evolution of cooperation* (ed. P. Hammerstein), pp. 7–36. Cambridge, MA: The MIT Press.

Fischbacher, U. 2007 z-Tree: Zurich toolbox for readymade economic experiments. *Exp. Econ.* **10**, 171–178. (doi:10.1007/s10683-006-9159-4)

Fischbacher, U. & Gächter, S. 2008 Heterogeneous social preferences and the dynamics of free riding in public goods. CeDEx discussion paper 2008-07, University of Nottingham.

Fischbacher, U., Gächter, S. & Fehr, E. 2001 Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* **71**, 397–404. (doi:10.1016/S0165-1765(01)00394-9)

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E. & Falk, A. 2007 Social comparison affects reward-related brain activity in the human ventral striatum. *Science* **318**, 1305–1308. (doi:10.1126/science.1145876)

Friedman, D. & Sunder, S. 1994 *Experimental methods. A primer for economists*. Cambridge, UK: Cambridge University Press.

Fudenberg, D. & Maskin, E. 1986 The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* **54**, 533–556. (doi:10.2307/1911307)

Gächter, S. 2007 Conditional cooperation: behavioral regularities from the lab and the field and their policy implications. In *Psychology and economics* (eds B. S. Frey & A. Stutzer), pp. 19–50. Cambridge, MA: The MIT Press.

Gächter, S. & Falk, A. 2002 Reputation and reciprocity: consequences for the labour relation. *Scand. J. Econ.* **104**, 1–26. (doi:10.1111/1467-9442.00269)

Gächter, S. & Fehr, E. 1999 Collective action as a social exchange. *J. Econ. Behav. Organ.* **39**, 341–369. (doi:10.1016/S0167-2681(99)00045-1)

Gächter, S. & Herrmann, B. 2005 Human cooperation from an economic perspective. In *Cooperation in primates and humans. Mechanisms and evolution* (eds P. M. Kappeler & C. P. van Schaik), pp. 267–289. Heidelberg, Germany: Springer.

Gächter, S. & Herrmann, B. 2007 The limits of self-governance when cooperators get punished—experimental evidence from urban and rural Russia. CeDEx discussion paper no. 2007-11, University of Nottingham.

Gächter, S. & Thöni, C. 2005 Social learning and voluntary cooperation among like-minded people. *J. Eur. Econ. Assoc.* **3**, 303–314. (doi:10.1162/154247605447323)

Gächter, S., Herrmann, B. & Thöni, C. 2004 Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence. *J. Econ. Behav. Organ.* **55**, 505–531. (doi:10.1016/j.jebo.2003.11.006)

Gächter, S., Renner, E. & Sefton, M. 2008 The long-run benefits of punishment. *Science* **322**, 1510. (doi:10.1126/science.1164744)

Gardner, A. & West, S. A. 2004 Cooperation and punishment, especially in humans. *Am. Nat.* **164**, 753–764. (doi:10.1086/425623)

Gintis, H. 2000 Strong reciprocity and human sociality. *J. Theor. Biol.* **206**, 169–179. (doi:10.1006/jtbi.2000.2111)

Gintis, H. 2008 Punishment and cooperation. *Science* **319**, 1345–1346. (doi:10.1126/science.1155333)

Goeree, J. K., Holt, C. A. & Laury, S. K. 2002 Private costs and public benefits: unraveling the effects of altruism and noisy behavior. *J. Public Econ.* **83**, 255–276. (doi:10.1016/S0047-2727(00)00160-2)

Gouldner, A. W. 1960 The norm of reciprocity: a preliminary statement. *Am. Sociol. Rev.* **25**, 161–178. (doi:10.2307/2092623)

Guala, F. 2005 *The methodology of experimental economics*. Cambridge, UK: Cambridge University Press.

Gunnthorsdottir, A., Houser, D. & McCabe, K. 2007 Disposition, history and contributions in public goods experiments. *J. Behav. Organ.* **62**, 304–315. (doi:10.1016/j.jebo.2005.03.008)

Gürerk, Ö., Irlenbusch, B. & Rockenbach, B. 2006 The competitive advantage of sanctioning institutions. *Science* **312**, 108–111. (doi:10.1126/science.1123633)

Hagen, E. H. & Hammerstein, P. 2006 Game theory and human evolution: a critique of some recent interpretations of experimental games. *Theor. Popul. Biol.* **69**, 339–348. (doi:10.1016/j.tpb.2005.09.005)

Haley, K. J. & Fessler, D. M. T. 2005 Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evol. Hum. Behav.* **26**, 245–256. (doi:10.1016/j.evolhumbehav.2005.01.002)

Hamilton, W. D. 1964 Genetical evolution of social behavior I, II. *J. Theor. Biol.* **7**, 1–52. (doi:10.1016/0022-5193(64)90038-4)

Hammerstein, P. (ed.) 2003 *Genetic and cultural evolution of cooperation*. Cambridge, MA: The MIT Press.

Hammerstein, P. & Hagen, E. H. 2005 The second wave of evolutionary economics in biology. *Trends Ecol. Evol.* **20**, 604–609. (doi:10.1016/j.tree.2005.07.012)

Hardin, G. 1968 The tragedy of the commons. *Science* **162**, 1243–1248. (doi:10.1126/science.162.3859.1243)

Henrich, J. 2004 Cultural group selection, coevolutionary processes and large-scale cooperation. *J. Econ. Behav. Organ.* **53**, 3–35. (doi:10.1016/S0167-2681(03)00094-5)

Henrich, J. & Boyd, R. 2001 Why people punish defectors. Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *J. Theor. Biol.* **208**, 79–89. (doi:10.1006/jtbi.2000.2202)

Henrich, J. & Henrich, N. 2007 *Why humans cooperate: a cultural and evolutionary explanation. Evolution and cognition series*. Oxford, UK: Oxford University Press.

Henrich, J. *et al*. 2005 'Economic man' in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behav. Brain Sci.* **28**, 795–815. (doi:10.1017/S0140525X05000142)

Henrich, J. *et al*. 2006 Costly punishment across human societies. *Science* **312**, 1767–1770. (doi:10.1126/science.1127333)

Herrmann, B. & Thöni, C. In press. Measuring conditional cooperation: a replication study in Russia. *Exp. Econ.* (doi:10.1007/s10683-008-9197-1)

Herrmann, B., Thöni, C. & Gächter, S. 2008 Antisocial punishment across societies. *Science* **319**, 1362–1367. (doi:10.1126/science.1153808)

Hix, S., Noury, A. & Roland, G. 2009 Voting patterns and alliance formation in the European Parliament. *Phil. Trans. R. Soc. B* **364**, 821–831. (doi:10.1098/rstb.2008.0263)

Hofstede, G. 2001 *Culture's consequences: comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: Sage.

Holm, H. J. & Nystedt, P. 2005 Intra-generational trust—a semi-experimental study of trust among different generations. *J. Econ. Behav. Organ.* **58**, 403–419. (doi:10.1016/j.jebo.2003.10.013)

Hopfensitz, A. & Reuben, E. In press. The importance of emotions for the effectiveness of social punishment. *Econ. J.*

Houser, D., Xiao, E., McCabe, K. & Smith, V. 2008 When punishment fails: research on sanctions, intentions and non-cooperation. *Games Econ. Behav.* **62**, 509–532. (doi:10.1016/j.geb.2007.05.001)

Inglehart, R. & Baker, W. E. 2000 Modernization, cultural change, and the persistence of traditional values. *Am. Sociol. Rev.* **65**, 19–51. (doi:10.2307/2657288)

Isaac, R. M. & Walker, J. M. 1988*a* Communication and free-riding behavior—the voluntary contribution mechanism. *Econ. Inq.* **26**, 585–608.

Isaac, R. M. & Walker, J. M. 1988*b* Group size effects in public goods provision: the voluntary contributions mechanism. *Q. J. Econ.* **103**, 179–199. (doi:10.2307/1882648)

Isaac, R. M., Walker, J. M. & Williams, A. W. 1994 Group-size and the voluntary provision of public goods—experimental evidence utilizing large groups. *J. Public Econ.* **54**, 1–36. (doi:10.1016/0047-2727(94)90068-X)

Kagel, J. & Roth, A. E. 1995 *The handbook of experimental economics*. Princeton, NJ: Princeton University Press.

Kelley, H. & Stahelski, A. 1970 Social interaction basis of cooperators' and competitors' beliefs about others. *J. Pers. Soc. Psychol.* **16**, 190–219.

Keser, C. & van Winden, F. 2000 Conditional cooperation and voluntary contributions to public goods. *Scand. J. Econ.* **102**, 23–39. (doi:10.1111/1467-9442.00182)

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V. & Fehr, E. 2006 Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* **314**, 829–832. (doi:10.1126/science.1129156)

Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J. & Sutter, M. 2008 Conditional cooperation on three continents. *Econ. Lett.* **101**, 175–178. (doi:10.1016/j.econlet.2008.07.015)

Kollock, P. 1998 Social dilemmas: the anatomy of co-operation. *Annu. Rev. Sociol.* **24**, 183–214. (doi:10.1146/annurev.soc.24.1.183)

Kreps, D., Milgrom, P., Roberts, J. & Wilson, R. 1982 Rational cooperation in the finitely repeated prisoners' dilemma. *J. Econ. Theor.* **27**, 245–252. (doi:10.1016/0022-0531(82)90029-1)

Krupp, D. B., Debruine, L. M. & Barclay, P. 2008 A cue of kinship promotes cooperation for the public good. *Evol. Hum. Behav.* **29**, 49–55. (doi:10.1016/j.evolhumbehav.2007.08.002)

Kurzban, R. & Houser, D. 2005 Experiments investigating cooperative types in humans: a complement to evolutionary theory and simulations. *Proc. Natl Acad. Sci. USA* **102**, 1803–1807. (doi:10.1073/pnas.0408759102)

Ledyard, J. O. 1995 Public goods: a survey of experimental research. In *The handbook of experimental economics* (eds A. E. Roth & J. H. Kagel), pp. 111–181. Princeton, NJ: Princeton University Press.

Lehmann, L. & Keller, L. 2006 The evolution of cooperation and altruism: a general framework and a classification of models. *J. Evol. Biol.* **19**, 1365–1376. (doi:10.1111/j.1420-9101.2006.01119.x)

Liebrand, W. B. G., Jansen, R. W. T. L., Rijken, V. M. & Suhre, C. J. M. 1986 Might over morality: social values and the perception of other players in experimental games. *J. Exp. Soc. Psychol.* **22**, 203–215. (doi:10.1016/0022-1031(86)90024-7)

Loewenstein, G., Thompson, L. & Bazerman, M. 1989 Social utility and decision making in interpersonal contexts. *J. Pers. Soc. Psychol.* **57**, 426–441. (doi:10.1037/0022-3514.57.3.426)

Madsen, E. A., Tunney, R. J., Fieldman, G., Plotkin, H. C., Dunbar, R. I. M., Richardson, J. M. & McFarland, D. 2007 Kinship and altruism: a cross-cultural experimental study. *Br. J. Psychol.* **98**, 339–359. (doi:10.1348/000712606X129213)

Marlowe, F. W. *et al.* 2008 More 'altruistic' punishment in larger societies. *Proc. R. Soc. B* **275**, 587–590. (doi:10.1098/rspb.2007.1517)

Marwell, G. & Ames, R. 1979 Experiments on the provision of public goods I: resources, interest, group size, and the free-rider problem. *Am. J. Sociol.* **84**, 1335–1360. (doi:10.1086/226937)

Masclet, D. & Villeval, M. C. 2008 Punishment, inequality, and welfare: a public good experiment. *Soc. Choice Welfare* **31**, 475–502. (doi:10.1007/s00355-007-0291-7)

Masclet, D., Noussair, C., Tucker, S. & Villeval, M. C. 2003 Monetary and nonmonetary punishment in the voluntary contributions mechanism. *Am. Econ. Rev.* **93**, 366–380.

Milinski, M. & Rockenbach, B. 2007 Spying on others evolves. *Science* **317**, 464–465. (doi:10.1126/science.1143918)

Milinski, M., Semmann, D. & Krambeck, H. J. 2002 Reputation helps solve the 'tragedy of the commons'. *Nature* **415**, 424–426. (doi:10.1038/415424a)

Mohan, N. H. 2008 Vengeance. NBER working paper no. 14131.

Monin, B. 2007 Holier than me? Threatening social comparison in the moral domain. *Int. Rev. Soc. Psychol.* **20**, 53–68.

Muller, L., Sefton, M., Steinberg, R. & Vesterlund, L. 2008 Strategic behavior and learning in repeated voluntary-contribution experiments. *J. Econ. Behav. Organ.* **67**, 782–793. (doi:10.1016/j.jebo.2007.09.001)

Neugebauer, T., Perote, J., Schmidt, U. & Loos, M. In press. Self-biased conditional cooperation: on the decline of cooperation in repeated public goods experiments. *J. Econ. Psychol.* (doi:10.1016/j.joep.2008.04.005)

Nikiforakis, N. 2008 Punishment and counter-punishment in public good games: can we really govern ourselves? *J. Public Econ.* **92**, 91–112. (doi:10.1016/j.jpubeco.2007.04.008)

Nikiforakis, N. & Normann, H. 2008 A comparative statics analysis of punishment in public goods experiments. *Exp. Econ.* **11**, 358–369. (doi:10.1007/s10683-007-9171-3)

Noussair, C. & Tucker, S. 2005 Combining monetary and social sanctions to promote cooperation. *Econ. Inq.* **43**, 649–660. (doi:10.1093/ei/cbi045)

Nowak, M. A. 2006 Five rules for the evolution of cooperation. *Science* **314**, 1560–1563. (doi:10.1126/science.1133755)

Nowak, M. A. & Sigmund, K. 2005 Evolution of indirect reciprocity. *Nature* **437**, 1291–1298. (doi:10.1038/nature04131)

Olson, M. 1965 *The logic of collective action*. Cambridge, MA: Harvard University Press.

Ones, U. & Putterman, L. 2007 The ecology of collective action: a public goods and sanctions experiment with controlled group formation. *J. Econ. Behav. Organ.* **62**, 495–521. (doi:10.1016/j.jebo.2005.04.018)

Oosterbeek, H., Sloof, R. & van de Kuilen, G. 2004 Cultural differences in ultimatum game experiments: evidence from a meta-analysis. *Exp. Econ.* **7**, 171–188. (doi:10.1023/B:EXEC.0000026978.14316.74)

Ostrom, E. 1998 A behavioral approach to the rational choice theory of collective action. *Am. Polit. Sci. Rev.* **92**, 1–22. (doi:10.2307/2585925)

Ostrom, E., Walker, J. M. & Gardner, R. 1992 Covenants with and without a sword—self-governance is possible. *Am. Polit. Sci. Rev.* **86**, 404–417. (doi:10.2307/1964229)

Page, T., Putterman, L. & Unel, B. 2005 Voluntary association in public goods experiments: reciprocity, mimicry, and efficiency. *Econ. J.* **115**, 1032–1052. (doi:10.1111/j.1468-0297.2005.01031.x)

Palfrey, T. R. 2008 Laboratory experiments. In *The Oxford handbook of political economy* (eds B. Weingast & D. Wittman), pp. 915–936. Oxford, UK: Oxford University Press.

Palfrey, T. R. & Prisbrey, J. E. 1997 Anomalous behavior in public goods experiments: how much and why? *Am. Econ. Rev.* **87**, 829–846.

Panchanathan, K. & Boyd, R. 2004 Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* **432**, 499–502. (doi:10.1038/nature02978)

Pillutla, M. & Murnighan, K. J. 1996 Unfairness, anger, and spite: emotional rejections of ultimatum offers. *Organ. Behav. Hum. Decis. Process.* **68**, 208–224. (doi:10.1006/obhd.1996.0100)

Rabin, M. 1993 Incorporating fairness into game-theory and economics. *Am. Econ. Rev.* **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.

Rege, M. & Telle, K. 2004 The impact of social approval and framing on cooperation in public good situations. *J. Public Econ.* **88**, 1625–1644. (doi:10.1016/S0047-2727(03)00021-5)

Reuben, E. & Riedl, A. In press. Public goods provision and sanctioning in privileged groups. *J. Conflict Resolut.* (doi:10.1177/0022002708322361)

Reuben, E. & van Winden, F. 2008 Social ties and coordination on negative reciprocity: the role of affect. *J. Public Econ.* **92**, 34–53. (doi:10.1016/j.jpubeco.2007.04.012)

Rockenbach, B. & Milinski, M. 2006 The efficient interaction of indirect reciprocity and costly punishment. *Nature* **444**, 718–723. (doi:10.1038/nature05229)

Sally, D. 1995 Conversation and cooperation in social dilemmas: a meta-analysis of experiments from 1958 to 1992. *Ration. Soc.* **7**, 58–92. (doi:10.1177/1043463195007001004)

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. 2003 The neural basis of economic decision-making in the ultimatum game. *Science* **300**, 1755–1758. (doi:10.1126/science.1082976)

Sefton, M., Shupp, R. & Walker, J. M. 2007 The effect of rewards and sanctions in provision of public goods. *Econ. Inq.* **45**, 671–690.

Seinen, I. & Schram, A. 2006 Social status and group norms: indirect reciprocity in a repeated helping experiment. *Eur. Econ. Rev.* **50**, 581–602. (doi:10.1016/j.euroecorev.2004.10.005)

Semmann, D., Krambeck, H. J. & Milinski, M. 2005 Reputation is valuable within and outside one's own social group. *Behav. Ecol. Sociobiol.* **57**, 611–616. (doi:10.1007/s00265-004-0885-3)

Seymour, B., Singer, T. & Dolan, R. 2007 The neurobiology of punishment. *Nat. Rev. Neurosci.* **8**, 300–311. (doi:10.1038/nrn2119)

Shinada, M. & Yamagishi, T. 2007 Punishing free riders: direct and indirect promotion of cooperation. *Evol. Hum. Behav.* **28**, 330–339. (doi:10.1016/j.evolhumbehav.2007.04.001)

Sigmund, K. 2007 Punish or perish? Retaliation and collaboration among humans. *Trends Ecol. Evol.* **22**, 593–600. (doi:10.1016/j.tree.2007.06.012)

Skyrms, B. 2009 Evolution of signalling systems with multiple senders and receivers. *Phil. Trans. R. Soc. B* **364**, 771–779. (doi:10.1098/rstb.2008.0258)

Sober, E. & Wilson, D. S. 1998 *Unto others. The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Sonnemans, J., Schram, A. & Offerman, T. 1999 Strategic behavior in public good games: when partners drift apart. *Econ. Lett.* **62**, 35–41. (doi:10.1016/S0165-1765(98)00203-1)

Sutter, M. 2007 Outcomes versus intentions: on the nature of fair behavior and its development with age. *J. Econ. Psychol.* **28**, 69–78. (doi:10.1016/j.joep.2006.09.001)

Sutter, M. & Kocher, M. G. 2007 Trust and trustworthiness across different age groups. *Games Econ. Behav.* **59**, 364–382. (doi:10.1016/j.geb.2006.07.006)

Sutter, M., Haigner, S. & Kocher, M. G. 2008 Choosing the carrot or the stick?—endogenous institutional choice in social dilemma situations. Working paper in economics and statistics 2008-07, University of Innsbruck.

Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. 2005 Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* **28**, 675–691. (doi:10.1017/S0140525X05000129)

Trivers, R. 1971 The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57. (doi:10.1086/406755)

van Prooijen, J. W., Gallucci, M. & Toeset, G. 2008 Procedural justice in punishment systems: inconsistent punishment procedures have detrimental effects on cooperation. *Br. J. Soc. Psychol.* **47**, 311–324. (doi:10.1348/014466607X218212)

Walker, J. M. & Halloran, M. A. 2004 Rewards and sanctions and the provision of public goods in one-shot settings. *Exp. Econ.* **7**, 235–247. (doi:10.1023/B:EXEC.0000040559.08652.51)

Weimann, J. 1994 Individual behavior in a free riding experiment. *J. Public Econ.* **54**, 185–200. (doi:10.1016/0047-2727(94)90059-0)

West, S. A., Griffin, A. S. & Gardner, A. 2007 Evolutionary explanations for cooperation. *Curr. Biol.* **17**, R661–R672. (doi:10.1016/j.cub.2007.06.004)

Wiessner, P. 2005 Norm enforcement among the Ju/'hoansi Bushmen—a case of strong reciprocity? *Hum. Nat. Interdiscip. Biosoc. Perspect.* **16**, 115–145.

Yamagishi, T. 1986 The provision of a sanctioning system as a public good. *J. Pers. Soc. Psychol.* **51**, 110–116. (doi:10.1037/0022-3514.51.1.110)

Zelmer, J. 2003 Linear public goods experiments: a meta-analysis. *Exp. Econ.* **6**, 299–310. (doi:10.1023/A:1026277420119)

Zizzo, D. J. 2003 Money burning and rank egalitarianism with random dictators. *Econ. Lett.* **81**, 263–266. (doi:10.1016/S0165-1765(03)00190-3)