University of Groningen

Put your money where your mouth is

Dijkstra, Jacob

Published in:
Rationality and Society

DOI:
[10.1177/1043463113492305](https://doi.org/10.1177/1043463113492305)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2013

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Dijkstra, J. (2013). Put your money where your mouth is: Reciprocity, social preferences, trust and contributions to public goods. *Rationality and Society*, 25(3), 290-334.
<https://doi.org/10.1177/1043463113492305>

Copyright

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

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Put your money where your mouth is: Reciprocity, social preferences, trust and contributions to public goods

Rationality and Society

25(3) 290–334

© The Author(s) 2013

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1043463113492305

rss.sagepub.com **Jacob Dijkstra**

University of Groningen/ICS, The Netherlands

Abstract

It is argued that trust and positive social preferences promote public goods production. However, public goods produced by any in-group may have favourable or unfavourable consequences for out-groups (called ‘benign’ and ‘malignant’ public goods, respectively). I develop a theoretical model of heterogeneous reciprocity preferences and report two experiments relating trust, social preferences and in-group bias to contributions to benign and malignant public goods. The results allow four general conclusions: (i) contributions to benign public goods are (weakly) higher than contributions to malignant ones; (ii) general trust is at best weakly related to contributions to both types of public goods; (iii) the expectation that others contribute is positively related to contributions to both types of public goods; and (iv) social preferences are positively related to contributions to benign public goods and unrelated to contributions to malignant public goods, while in-group bias is negatively related to contributions to both public goods.

Keywords

In-group bias, public goods, reciprocity, social preferences, trust

Corresponding author:

Jacob Dijkstra, Department of Sociology, University of Groningen/ICS, Grote Rozenstraat 31, 9712 TG Groningen, The Netherlands.

Email: j.dijkstra@rug.nl

Introduction

Many studies from different branches of the social sciences show that social capital is related to beneficial collective outcomes, such as economic growth (Knack and Keefer, 1997), increased health (e.g. Fujiwara and Kawachi, 2008; Kawachi et al., 1997), less criminal violence (e.g. Galea et al., 2002), enhanced academic performance (e.g. Coleman, 1988) and more efficient community resource management (Bouma et al., 2008), among other things. Although there is no consensus of what the definition of social capital should be, ‘[t]rust, cooperative norms, and associations within groups each fall within the elastic definitions that most scholars have applied to the term social capital’ (Knack and Keefer, 1997: 1251–1252, emphasis added). Moreover, social capital is generally treated as ‘characteristic of communities (...)’, and described in terms of ‘(...) *trust*, norms and networks that enable *collective action*’ (Bouma et al., 2008: 155, emphases added).

A central theoretical mechanism underlying the relation between social capital and the advantageous outcomes mentioned above is that groups with a lot of social capital are assumed to be better able to solve collective action problems. Thus, according to Putnam (1995: 66, emphases added) ‘(...) “social capital” refers to features of social organization such as networks, norms, and *social trust* that facilitate *coordination and cooperation for mutual benefit*.’

The present paper focuses on the link between individuals’ trust in others and collective action. I define trust as an individual’s beliefs about trustworthy behaviour of either general or specific others (both forms of trust are investigated in this study). I then pose the following two research questions: (i) do individuals with high levels of trust actually contribute more to public goods than other individuals, and (ii) do they do so regardless of whether the production of the public good has positive or negative consequences for the larger society? The latter research question pertains to the issue that a public good produced by a certain (in-)group might have positive or negative consequences for one or more out-groups.

Assuming that the positive relation between trust and contributions to public goods exists, the question arises as to which theoretical mechanism underlies this relation. Note that this theoretical mechanism is not obvious: why would any rationally egoistic individual contribute to a public good when she trusts others will, since from the perspective of a rational egoist not contributing is the best response, no matter what others do? In this paper I argue that a theoretical model based on *heterogeneous reciprocity preferences* provides the theoretical link between trust and contributions to public goods.

The second research question above recognizes that public goods produced by any particular group of individuals might not be beneficial from the point of view of certain out-groups or the entire society. Although a public good for the group producing it, these goods might constitute a 'public bad' for the out-group(s) or for society. In this respect, Olson (1982) discusses rent-seeking interest groups lobbying for policies that are beneficial for group members, but impose large costs on society as a whole. Also relevant to this issue is Putnam's distinction between *bonding* and *bridging* social capital (Putnam, 2007), being related to in-group and out-group solidarity, respectively. Thus, while a lot of bonding social capital might be beneficial for the in-group, it might well harm out-group members.

The answers to the research questions determine how we should evaluate observed levels of social trust from a societal perspective. If trust is positively related to contributions to public goods, *regardless* of whether these public goods have positive or negative externalities for out-groups in society, we cannot generally say that societies with high levels of trust are 'better off' than other societies. For if such a 'high-trust' society were characterized by the existence of many separate groups producing public goods with negative externalities for each other, the overall result would likely be worse than would have been the case had trust been low. In the words of Bornstein and Ben-Yossef (1994: 65), '(...) cooperation is a good thing by definition only in the limited context of a one-group social dilemma (...). In intergroup conflicts (...) universal cooperation is collectively deficient.' Intergroup conflicts are described by these authors as public goods games where the production of the public good by the in-group has negative consequences for the out-group, which maps precisely onto the second research question of this paper. Note also how the second research question harkens back to the distinction made by Hechter et al. (1992) between 'local' and 'global' order.

The following section describes the theoretical mechanism linking trust to contributions to public goods. This section is followed by the report of two experimental studies in which hypotheses derived from the theory are tested. The description of each study contains the hypotheses tested, the methods used, the results of the experiment, and a brief discussion. The paper is concluded by a general discussion of the results and their implications.

Theory

The public goods game and rational egoism

I start by describing a general public goods game, without first making any in-group/out-group distinction. Let there be I players in the game. Each

player i has an initial endowment e_i , which is assumed to be equal for all players. At the start of the game each player makes one decision as to how much of her endowment to invest in the public good, after which the game ends. Denote the investment by x_i . No player knows the investment decision of any other player when making her own decision. Let c be the rate of return of the investment in the public good, with $1 < c < 1$. The payoff function u_i of player i is

$$u_i = \frac{c \sum_{j \in I} x_j}{I} + (e_i - x_i) \quad (1)$$

Thus, a player's payoff consists of two parts: (i) an even share of the public good produced by the group (first term after the = sign); and (ii) the part of her initial endowment she did not invest (second term after the = sign).

To determine what a rational player would do in this game we need to make assumptions concerning her preferences. Below I expound a theoretical model based on heterogeneous reciprocity preferences. Such a model is a member of the class of social preference models (e.g. Fehr and Fischbacher, 2002) that can be contrasted with the traditional rational choice postulate of rational egoism. According to this postulate, each player is better off not investing anything in the public good no matter what the other players decide. To see this, note that if a player increases her investment with one unit, she increases the first term of (1) by $\frac{c}{I}$, which is less than 1. The second term of (1), however, is decreased by 1.

Thus, based on the rational egoism postulate the prediction is that each player i earns exactly e_i in this game. However, had all players invested their entire endowment in the public good each would have earned ce_i and would thus have been better off. Hence, we have a *social dilemma* (e.g. Dawes, 1980): individually rational and egoistic behaviour leads to collectively undesirable outcomes. Before explaining the theoretical model based on reciprocity preferences, I first briefly introduce two types of public goods and the two ideal-typical societies that will be investigated in the experiments of Study 1 and Study 2.

Harmony and Conflict

Consider the effects of the public good introduced above on the larger society. What I call 'benign' public goods, are goods whose production is not only beneficial for the members of the in-group producing it, but also for other individuals in society who are not members of the particular in-group.

An example would be fund-raising activities of a charity organization. Such activities frequently resemble the production of public goods in that the contribution of any individual member has relatively little impact on the aggregated outcome of the activity (i.e. the total sum of money raised), whereas participating in the activity entails considerable costs for the individual (e.g. in terms of time and effort spent soliciting and collecting donations). Production of this public good, however, has positive consequences for the beneficiaries of the charity.

'Malignant' public goods, on the other hand, are public goods whose production is beneficial for the members of the producing in-group, but harmful for out-group members in society. Activities of criminal organizations such as street gangs provide examples. Many of their activities (such as warfare against a rival gang) share important characteristics with public goods. For instance, every individual gang member wants his own gang to win the gang war, but prefers to let his fellow members do the actual fighting. Production of this public good has negative consequences for out-group members (both members of the rival gang and members of the general public).

Now consider two types of societies, and call them *Harmony* and *Conflict*, respectively. In *Harmony* all public goods produced are of the benign type, whereas in *Conflict* all public goods are malignant. *Harmony* and *Conflict* are extreme cases, since real-life societies typically exhibit a mix of benign and malignant public goods, but they serve the theoretical exposition in this paper well and will be investigated experimentally.

Reciprocity preferences

In opposition to rational egoism, a large amount of research, especially in behavioural economics and psychology, shows that many individuals have *social preferences*: they are not only motivated by their own material interests, but also care (either positively or negatively) about the material interests of others (e.g. Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999). A very important type of social preference is the *preference for reciprocity*, according to which individuals reward other players' kind intentions or behaviours and punish their unkind intentions or behaviours (e.g. Falk and Fischbacher, 2006; Fehr and Gächter, 2002; Rabin, 1993, 2002). These reciprocity preferences are distinct from so-called 'repeated game incentives', according to which a participant behaves in a reciprocal fashion only to get rewarded for this at a later time. From the perspective of this paper, a particularly instructive theory of reciprocal behaviour is the social exchange heuristic of Kiyonari et al. (2000) and Yamagishi et al. (2007).

Table 1. A Prisoner's Dilemma Game (PDG) and an Assurance Game (AG); adapted from Kiyonari et al. (2000).

		PDG		AG	
		Player A		Player A	
		Cooperate	Defect	Cooperate	Defect
Player B	Cooperate	2, 2	0, 3	2, 2	0, 1
	Defect	3, 0	1, 1	1, 0	1, 1

Note: The first entry in each cell indicates payoff of player B; the second entry indicates payoff of player A.

The central claim of the social exchange heuristic is that people generally have a cognitive bias in the information processing of social exchange, in the sense that they tend to perceive a Prisoner's Dilemma Game (PDG) as if it were an Assurance Game (AG).

Consider the PDG of Table 1. It can be regarded as a discrete, two-player version of the general public goods game described earlier. Thus, assuming players A and B make a one-time simultaneous decision to choose either Cooperate or Defect, each is guided by her self-interest to choose Defect, since this is her best choice regardless of the choice of the other player. This individually rational course of action yields them both a payoff of 1. Had they both chosen Cooperate, however, they each would have gotten the larger payoff of 2. Note how in the PDG 'trust', in the sense of the expectation of the behaviour of the other player, plays no role in the determination of an individual's rational action. It is best to play Defect regardless of what the other chooses, so trusting that the other will play Cooperate should have no effect on one's choice.

Kiyonari et al. (2000), however, claim that many people actually perceive this PDG as an AG, also depicted in Table 1. Note how in the AG a player wants to play Cooperate provided the other player also plays Cooperate. If the other chooses Defect, however, a player will prefer to play Defect, too. Thus, contrary to the PDG, in the AG trust does play an important role in determining an individual's rational course of action. If she trusts the other to play Cooperate, an individual should also play Cooperate; if she does not trust the other player, she should play Defect.

Kiyonari et al. (2000) cite evidence showing that many subjects in PDG experiments indeed rate the desirability of the outcome of mutual cooperation higher than that of the outcome resulting from defecting on a cooperating partner. Similarly, Rilling et al. (2002) offer questionnaire and neurological

evidence from an experiment using the PDG of Table 1 (with payoffs being dollar amounts), showing that mutual cooperation is the most desirable outcome for participants, and is associated with consistent activation in brain areas related to reward processing. Finally, the social exchange heuristic is also consistent with the account of Frank (1988) concerning the 'impulse control problem': humans excessively (and hyperbolically) discount future payoffs compared to current ones, making enduring cooperation hard to sustain, since individuals are not susceptible to the threat of future losses due to a breakdown of cooperation. The mental transformation of a PDG into an AG causes individuals to actually prefer mutual cooperation, thus dissipating the impulse control problem.

The transformation of a PDG into an AG proposed by the social exchange heuristic implies reciprocal behaviour: individuals want to repay both kind behaviour with kind behaviour (i.e. cooperate when the other player cooperates), and unkind behaviour with unkind behaviour (i.e. defect when the other player defects). The seminal work on public good games by Fehr and Gächter (2002) suggests that reciprocal preferences also play an important role in the multi-player public goods games of the current paper. Moreover, Fehr and Gintis distinguish reciprocity preferences from the previously mentioned repeated game incentives by stating that (2007: 49) '(...) reciprocity is not simply long-term, enlightened self-interest' to build a good reputation, but is also practiced when an actor's reciprocal acts '(...) obviously reduce his or her economic net gain.' Reciprocity preferences in turn allow trust to play a role in the individual's decision making.

Heterogeneous reciprocity preferences

Application of a theoretical model of reciprocity preferences to Harmony and Conflict raises the issue that an individual player can take one of two perspectives. She can either take an 'in-group perspective', caring mainly about the payoffs her fellow group members receive, or she can take a 'society perspective', caring equally about the payoffs of all members of society.

Research in psychology and economics suggests that there is heterogeneity in the social preferences people have. Reviewing the behavioural economics literature on this point, Fehr and Gintis (2007) point to the co-existence of individuals having strong reciprocity preferences and purely egoistic individuals. Similarly, Van Lange et al. (2007) distinguish between three types of social preferences (prosociality, individualism and competitiveness) that are consistently found in the population (see Balliet et al. (2009) and Bogaert et al. (2008) for recent overviews of this *social value*

orientation research and its relation to cooperation in social dilemmas). Specifically with respect to the social exchange heuristic, Simpson (2004) shows that part of the experimental participants apply the transformation predicted by the heuristic and play the PDG as if it were an AG, while others play the PDG 'as is'.

With respect to situations in which multiple groups exist, there is evidence that individuals favour their own group members over members of the out-groups. This phenomenon is called *in-group favouritism* or *in-group bias*. The 'minimal-group' paradigm (e.g. Brown, 1986; Tajfel and Turner, 1979) contends that even seemingly trivial distinctions between the in-group and out-group can spark in-group favouritism. Game-theoretic and computational analyses of Choi and Bowles (2007) show that a preference for 'parochial altruism', which is essentially a combination of in-group favouritism and hostility towards out-group members, could indeed have evolved. However, the extent to which 'in-group love' is necessarily mirrored by 'out-group hate' is a debated issue (Allport, 1954; Brewer, 1999).

In light of this previous research I propose the assumption of 'heterogeneity in reciprocity preferences'. Thus, individuals differ in the strength of their reciprocity preferences, with some being purely selfish. In line with the social exchange heuristic I assume that reciprocal players will want to contribute to the public good if they trust others to do so. If they do not expect others to contribute, they make the choice that maximizes their own payoff. In addition to these heterogeneity in general social preferences, I assume there is heterogeneity in in-group bias: of those having reciprocity preferences, some will exhibit a strong bias in favour of in-group members, and others will not. I now apply this theoretical model of heterogeneous reciprocity preferences to the public goods in Harmony and Conflict.

Heterogeneous reciprocity preferences in Harmony and Conflict

Harmony and Conflict are like the general public goods game but for one characteristic: public goods produced by an in-group have positive (Harmony) or negative (Conflict) consequences for out-groups in society. Let G be the set of groups in society with $|G|$ being the number, and let G_i denote the group of which actor i is a member, with $|G_i|$ being the number of its members. Assume for simplicity that all groups have an equal number of members and that a unit of contribution in a particular in-group has the same impact a on the value of the public goods of all out-groups. The payoffs for all players i are then assumed to be

$$u_i = \frac{c \sum_{j \in G_i} x_j + a \sum_{g \in G, g \neq G_i} \sum_{k \in g} x_k}{|G_i|} + (e_i - x_i) \quad (2)$$

where the parameter c is now restricted to $1 < c < |G_i|$.

Equation (2) shows that actor i 's payoff is a function of the contributions of her fellow in-group members $j \in G_i$, just as in Equation (1) where we had only one group. In addition, the contributions of out-group members in the rest of society (summed within each out-group, and then summed over all out-groups) have an impact a on the value of the public good produced by group G_i . Note how the resulting value of the public good of group G_i is shared among the members of G_i only.

The case of Harmony is now defined by $a > 0$, while in Conflict we have $a < 0$. In addition, I assume that in Conflict $a < \frac{1-c}{|G|-1}$. To find the level of contribution a selfish player i would make, again consider what happens to (2) if i increases her investment in the public good with one unit. Her payoffs are increased by $\frac{c}{|G_i|} < 1$, because of the extra units of public good produced by her in-group, but are decreased by 1 because of her investment. Thus, *individual rationality* dictates a purely selfish player i to contribute nothing to her in-group's public good, regardless of what out-groups do and regardless of whether players are in Harmony or in Conflict. If all players from all groups follow this dictate, each earns e_i .

Group rationality, however, would dictate otherwise. Assume members of G_i all contribute their entire endowment. The payoffs for all members of

$$G_i \text{ would then be } ce_i + \frac{a \sum_{g \in G, g \neq G_i} \sum_{k \in g} x_k}{|G_i|} > e_i + \frac{a \sum_{g \in G, g \neq G_i} \sum_{k \in g} x_k}{|G_i|}, \text{ since } c > 1.$$

Thus, within any group all members are better off if they all contribute their entire endowment to their in-group's public good, compared to the situation in which they contribute nothing. Note that this is true regardless of the contributions of members of out-groups to *their* public goods (expressed by

$$\text{the term } \frac{a \sum_{g \in G, g \neq G_i} \sum_{k \in g} x_k}{|G_i|}, \text{ and regardless of whether players are in Harmony or in Conflict (i.e. regardless of the value of } a).$$

The difference between Harmony and Conflict arises at the level of *societal rationality*, that is, the level of all groups taken together. First

Table 2. Contributions to the group public good as prescribed by different levels of rationality in Harmony and Conflict.

Level of rationality	Type of society	
	Harmony	Conflict
Individual rationality	Contribute nothing	Contribute nothing
Group rationality	Contribute all	Contribute all
Societal rationality	Contribute all	Contribute nothing

Note: Since the payoff function (2) is linear in the contributions, different levels of rationality prescribe to contribute either all or nothing of a player's endowment.

consider Harmony, and assume all players in society contribute their entire endowment to their own group's public good. In this case any player i earns $[c + a(|G| - 1)]e_i > e_i$. Thus, a player earns ce_i from his own group and $a(|G| - 1)e_i > 0$ from all other $|G| - 1$ groups. Since this is more than what each player would earn when all would follow individual rationality (e_i), *societal rationality dictates players to contribute their entire endowment in Harmony*. In Conflict, however, $[c + a(|G| - 1)]e_i < e_i$, since $a < \frac{1-c}{|G|-1} < 0$. Thus, if all players would contribute their entire endowment in Conflict, all would be worse off than if nobody had contributed anything, and *societal rationality dictates players to contribute nothing in Conflict*. Table 2 summarizes this analysis.

Let us now apply the heterogeneous reciprocity preferences to Harmony and Conflict. As mentioned above, selfish players always follow the dictates of individual rationality, and thus never contribute. Reciprocal players with a strong in-group bias take a group perspective, and will contribute to their in-group's public good when they trust other members in their in-group will contribute. When they do not trust others to contribute, they will themselves also withhold their contribution. Moreover, they behave this way both in Harmony and Conflict.

Reciprocal players who take a society perspective, that is, who do not have a (strong) in-group bias, never contribute in Conflict, regardless of what they expect others to do. To see this, note that in Conflict the payoffs for all members of society are best served when no one contributes to his or her in-group's public good. Thus, from the societal perspective, the group public goods are malignant and constitute 'public bads'. From the society perspective, contribution to the 'true, societal' public good means refraining from contributing to the public good

Table 3. Decisions to contribute to the group public good for different player types in Harmony and Conflict.

Player type	Trust	Harmony	Conflict
Reciprocal	Yes	Contribute	Contribute
Group perspective	No	Not contribute	Not contribute
Reciprocal	Yes	Contribute	Not contribute ^a
Society perspective	No	Not contribute	Not contribute ^a

^aFor Reciprocal Society Perspective players in Conflict, trust means they expect other players *not* to contribute to their group public goods.

of one’s in-group. Thus, when reciprocal players without a strong in-group bias trust others will *not* contribute to their group public goods, they themselves do not contribute to the public good of their in-group. However, when they expect that others *will* contribute to their respective group public goods, reciprocal players without a strong in-group bias will still *not* contribute to their group public good, since this is the best they can do for themselves given that others cannot be trusted to withhold their contributions.

Finally, in Harmony, contributing to the public good of one’s in-group increases the payoffs for all members of society, inducing reciprocal players without a strong in-group bias to contribute to their in-group public good when they trust others will, and to withhold their contribution to their in-group public good when they do not trust others to contribute. This analysis is recapitulated in Table 3 below. Since selfish players never contribute, I leave them out.

Overview of studies

Study 1 implements the Harmony and Conflict societies in a two-group setting. Central to Study 1 is how general and specific trust relate to contributions in both types of societies. General trust is measured using different sets of questionnaire items. Social preferences and in-group bias are left unmeasured in Study 1. Study 2 tries to replicate the results of Study 1, with regard to the effects of general and specific trust. In addition, social preferences and in-group bias are measured and their impact on contributions in Harmony and Conflict analysed. The description of each study consists of the hypotheses tested, the methods used, the results and a brief discussion.

Study 1

Hypotheses. The last two columns of Table 3 show that in Harmony both types of reciprocal players contribute when they trust others to contribute. In Conflict, however, only trusting players taking the group perspective contribute. Thus, I formulate Hypothesis 1.¹

H1: Contribution to the in-group public good is higher in Harmony than in Conflict.

H1 is in contradiction with results of the ‘minimal-group’ paradigm (Brown, 1986; Tajfel and Turner, 1979), and with the team game experiments of Bornstein and colleagues (e.g. Bornstein, 1992, 2003; Bornstein and Ben-Yossef, 1994; Bornstein et al., 2008). Both traditions suggest that individuals contribute *more* to their in-group’s public good in a situation like Conflict than in a situation like Harmony. Both approaches assume individuals identify with their in-group and fellow in-group members, which is assumed to spark a competitive attitude vis-à-vis the out-groups. The current model differs from this in two important respects: (i) in line with psychological research (e.g. Brewer, 1999) it does not a priori assume any kind of ‘out-group hate’ or competitiveness; and (ii) it assumes heterogeneous reciprocity preferences, with part of the reciprocal players taking a group perspective, while others take a society perspective. These two differences cause H1 to deviate from the minimal-group prediction.

Note that groups in the current paper are truly ‘minimal’ in the sense that they are solely defined by the payoff structure of a *one-shot* game. However, if for instance the game were played repeatedly such that a common group history of play developed, H1 would not preclude the possibility that group identification would then be strengthened to the degree that contributions in Conflict would exceed those in Harmony.

The rows in Table 3 indicate that not trusting others to contribute induces all reciprocal player types not to contribute themselves. Trusting others, however, makes some player types contribute to their in-group’s public good. Hence, Hypothesis 2 is:

H2: Contribution to the in-group public good is positively related to the expectation that others contribute.

H2 is consonant with experimental evidence from social psychology. Mulder et al. (2006: 148), reviewing the literature on this point, state that ‘(...) *trust* that others will cooperate is an important determinant of cooperation (...)’.

The cells in Table 3 show that trusting others to contribute to the group public good makes both reciprocity player types contribute in Harmony, whereas it makes only those among them that take the group perspective contribute in Conflict. This yields Hypothesis 3:

H3: The relation between contribution to the in-group public good and the expectation that others contribute to it, is stronger in Harmony than in Conflict.

Hypotheses on general trust and specific trust

The trust of Hypotheses 1–3 is defined in two distinct ways. *General trust* is an individual's beliefs about trustworthy behaviour of general others. Anderson et al. (2004) found that the most frequently employed measure of general trust (the response to the question whether 'most people can be trusted') is positively related to contributions in a public goods experiment. Glaeser et al. (2000) found that measures of general trust do not predict *trust placed* in an experimental trust game as much as *being trustworthy* in such a game. For *specific trust* I employ the definition from experimental trust game research, where '(...) trust is the willingness to bet that another person will reciprocate a risky move (at a cost to themselves)' (Camerer, 2003: 85). The risky move in this case is contributing to a public good.

As argued in the introduction, if general trust is a component of social capital, it should be positively related to contributions to the in-group public goods in Harmony and negatively related to such contributions in Conflict. If this were the case, a society whose members have much general trust could genuinely be said to possess social capital in the sense that only benign public goods would be produced and not malignant ones. This yields Hypothesis 4.

H4: General trust is positively related to contributions to the in-group public goods in Harmony and negatively related to such contributions in Conflict.

Since specific trust is situation specific, I expect that the effect of general trust on contributions is mediated by specific trust, yielding Hypothesis 5.

H5: The effect of general trust on contribution to the in-group public goods is mediated by specific trust in both Harmony and Conflict.

Methods

Design

Each participant of the experiment of Study 1 was randomly assigned to one of two groups of three participants each, and started out with an endowment of 10 points. The rate of return of the public good was $c = 2$, while the effect of contributions of members of the other group was $a = \frac{3}{2}$ in Harmony, and $a = -\frac{3}{2}$ in Conflict.²

I used a within-subjects design with each participant completing the following three experimental tasks. Task 1 was the completion of a questionnaire measuring general trust, along with some background variables. Task 2 asked the participant to specify the number of points s/he wanted to contribute to his or her in-group's public good in Harmony (Task 2a), after which the participant was asked to indicate how much s/he expected other participants (both fellow group members and members of the other group) had contributed on average (Task 2b). Task 3 was identical to Task 2, but then for Conflict. Participants were free to contribute any number between 0 and 10 (inclusive).

The Harmony and Conflict situations were one-shot games, that is, Tasks 2 and 3 were completed only once by each participant. Participants received no feedback about choices of other participants during or after the experiment, to prevent spill-over effects between Tasks 2 and 3. The lack of information about choices of others makes the games effectively simultaneous, as was assumed in the theory discussed in the previous section.

One point scored in the experiment was worth 25 eurocents. In Harmony and Conflict participants earned points according to Equation (2). In addition, participants earned an additional 10 points each time their guess of other participants' average contributions was at most 2 points from the actual average contributed by the others. Since participants had to make four such guesses (one for fellow in-group members and one for out-group members, in both Harmony and Conflict), they could earn up to 40 additional points with this. By paying participants for their assessments of how much others contributed, I tried to induce participants to honestly give their best estimate.

Participants were given feedback about the points they earned in the experiment only after all tasks had been completed by everyone. They learned only the sum of points they had earned in the entire experiment, and not whether they had earned more in Task 2 or in Task 3, or whether they had earned more through their contribution decisions (Tasks 2a and 3a) or through their guessing of the contributions of others (Tasks 2b and 3b). In this way participants were unable to reconstruct who had been their fellow in-group members or who had been in the out-group, even after receiving payments.

Table 4. The order of tasks in experimental sessions in Study 1; the number of participants is given in brackets.

	Task 2 before Task 3	Task 3 before Task 2
Task 1 first	2 (18)	2 (24)
Task 1 last	2 (12)	1 (6)

An experimental session consisted of the tasks described above followed by the payment of the participants. Seven sessions were organized in which either 6 or 12 participants took part. The number of participants present had neither an effect on the contribution decisions of participants nor on their assessment of contributions of others. The order of the tasks was randomized, as shown in Table 4. The unbalanced nature of Table 4 is due to participants not showing up.

Participants

Sixty participants (35 female and 25 male with an average age of about 21, ranging from 18 to 24) participated in the experiment. They were recruited among the undergraduates of a social science faculty of a large university through an email emphasizing the possibility of earning cash. The experiment was also advertised on the central information screen in the faculty’s cafeteria. In all recruitment efforts and in the experimental instructions I emphasized that the experiment did not involve any deception.

Materials

The experiment was programmed in Z-Tree developed by Fischbacher (2007), and the entire experiment was run on the computer. Participants’ estimates of the contributions made by others, in Tasks 2b and 3b, constitute the measurements of specific trust.

I used several existing scales to measure general trust. The Interpersonal Trust Scale (ITS; Rotter, 1967, 1971) contains 25 items asking participants to what extent they agree with statements such as: ‘Fear and social disgrace or punishment rather than conscience prevents most people from breaking the law’. The ITS can be found in Robinson et al. (1991) and is also used by Glaeser et al. (2000). I also included the well-known General Social Survey (GSS) general trust item: ‘Generally speaking, would you say that most people can be trusted or that you can’t be too careful in life?’ In the remainder, I refer to it as the GSS trust question.

In addition, I used the Faith in People Scale (Rosenberg, 1957), the Attitudinal Measures of Trust, the Behavioral Measures of Trust and the Voluntary Participation Scale (Anderson et al., 2004). Finally, I used two items from Glaeser et al. (2000) that were not already contained in the previous scales. With a Cronbach's alpha of 0.73, only the ITS had acceptable reliability: all others had alphas of 0.52 or lower. The low reliability of the scales most likely is due to the relatively homogeneous experimental sample. As measures of general trust I therefore only retain the ITS and the GSS trust question for the analyses.

I translated all the scales mentioned above from English to Dutch. ITS scale scores for each participant were computed by adding the item scores. In Appendix 1 you will find the ITS and the GSS trust question. The background variables I measured in Task 1 were sex and age. In addition, I asked participants how many of the others present they considered to be their personal friends.

Procedure

Participants were assigned to a computer terminal by the experiment leader. They were seated in a large computer room in such a way as to prevent them from reading each other's screens. The computer randomly assigned participants to one of two (in sessions with six participants) or four (in sessions with 12 participants) groups. In sessions with four groups, the computer also randomly made two pairs of groups. Neither the participants nor the experiment leader knew the group composition.

Participants received instructions on their computer screens, in addition to a handout explaining Tasks 2 and 3 (see Appendix 2 for the handout pertaining to the Harmony condition). Whenever participants had any questions they could raise their hand and the experiment leader would come to give extra explanation. This never occurred, however.

All participants in a session completed the tasks in the same order, and had to wait for all others before progressing to the next task. After all participants had completed the final task, everyone was shown their private earnings, which could be collected from the experiment leader. A session lasted for about 25 minutes and participants earned an approximate average of 8.30 Euros.

Results of Study I

Description of preference types

A full categorization of participants in terms of selfish preferences and the reciprocal preference types of Table 3 would require data on participants'

contributions given different levels of expected contributions of others (see Fischbacher et al., 2001). While not containing this kind of information, the data from the current experiment do provide us with a first indication of heterogeneity in preferences.

The average observed contributions in Harmony and Conflict were 6.08 ($SD = 3.47$) and 5.35 ($SD = 3.81$), respectively. Of the 60 participants, there were only two truly selfish participants who contributed nothing in both Harmony and Conflict. To get an idea of heterogeneity in reciprocal preferences, I compared participants' own contributions to their estimates of the contributions of others. A contribution that is higher than the expected contributions of others is labelled 'overinvestment'. A contribution that is lower than the expected contributions of others is labelled 'underinvestment'.

There were five participants who overinvested with respect to their expectations concerning both fellow in-group members and members of the out-group in Harmony, and who underinvested with respect to their expectations concerning both groups in Conflict. Roughly speaking, these participants can be labelled 'reciprocal with a society perspective'. There were 16 participants who overinvested with respect to their expectations concerning their in-group in both Harmony and Conflict. Roughly speaking, they can be labelled 'reciprocal with a group perspective'. Again, because of the one-expectation-one-decision data, this classification is only approximate. In addition, it is incomplete: only 23 participants out of 60 can be classified.

Other descriptive results

To shed some first light on the main relations of interest in Study 1, Figures 1 and 2 show the scatter plots of participants' contributions in Harmony and Conflict related to their trust in others. Since the GSS trust question is a dichotomous item, I omit it. The figures show that in both Harmony and Conflict specific trust is strongly related to contributions, but general trust (the ITS) is not. The two specific trust measures (related to in-group and out-group members) are strongly related in both conditions, but seem unrelated to the ITS. To explore these relations more formally, I now turn to the hypotheses testing.

Hypothesis testing

The mean contribution of 6.08 ($SD = 3.47$) in Harmony is slightly higher than the mean contribution of 5.35 ($SD = 3.81$) in Conflict, as predicted by H1, and contradicting the alternative minimal-group hypothesis. However, this difference is only marginally significant (paired sample

Figure 2. Scatterplot matrix main variables in Study 1, Conflict: ConConflict: contributions in Conflict; ITS: Interpersonal Trust Scale; SpecTrustINC: specific trust in in-group members in Conflict; SpecTrustOUTC: specific trust in out-group members in Conflict.

I test Hypotheses 2, 4 and 5 by conducting mediation analyses using ordinary least squares (OLS) regression. Participants’ estimates of the contributions by in-group members and out-group members were highly correlated (Pearson correlations of 0.87 and 0.85, $p < 0.001$ in Harmony and Conflict, respectively), as suggested by Figures 1 and 2. For the analyses of Tables 5 and 6, I therefore constructed a single variable, Specific Trust, measuring the participants’ estimates of the contribution of an average other participant, irrespective of in-group or out-group membership. The means of Specific Trust in Harmony and Conflict were 5.40 (SD = 2.47) and 4.95 (SD = 2.45), respectively.

The two mediation analyses (one for Harmony and one for Conflict) have the same structure. In Model I the variable Specific Trust is regressed

Table 5. Mediation analysis using ordinary least squares regression for Harmony in Study I; standard errors are given in brackets.

Dependent →	Model I	Model IIa	Model IIb	Model III	Model IVa	Model IVb	Model Va	Model Vb
	Specific Trust	Specific Trust	Specific Trust	Contribution	Contribution	Contribution	Contribution	Contribution
Constant	1.76 (5.10)	-2.51 (5.59)	1.76 (5.20)	-0.74 (6.90)	-9.49 (7.28)	-0.91 (7.02)	-6.91 (4.50)	-2.82 (4.24)
# Friends	0.10 (0.41)	0.07 (0.40)	0.10 (0.43)	0.81 (0.56)	0.74 (0.53)	0.83 (0.58)	0.66 ^{***} (0.33)	0.72 ^{***} (0.35)
Sex	1.20 [*] (0.67)	1.04 (0.66)	1.20 [*] (0.68)	1.86 ^{***} (0.90)	1.52 [*] (0.86)	1.88 ^{***} (0.91)	0.46 (0.54)	0.58 (0.57)
Age	0.15 (0.24)	0.11 (0.24)	0.15 (0.24)	0.23 (0.32)	0.15 (0.31)	0.23 (0.33)	0.03 (0.19)	0.07 (0.20)
Harmony First	-0.05 (0.66)	-0.18 (0.66)	-0.05 (0.67)	1.13 (0.90)	0.87 (0.85)	1.12 (0.91)	1.05 ^{**} (0.53)	1.17 ^{***} (0.55)
Experiment First	-0.27 (0.79)	-0.20 (0.77)	-0.27 (0.80)	0.02 (1.06)	0.16 (1.01)	0.04 (1.08)	0.37 (0.62)	0.33 (0.65)
ITS		0.07 ^{**} (0.04)			0.15 ^{***} (0.06)		0.08 ^{**} (0.04)	
GSS trust			-0.004 (0.81)			0.21 (1.09)		0.21 (0.67)
Specific Trust							1.03 ^{***} (0.11)	1.08 ^{***} (0.11)
Δ R ² (p)	0.07 (0.55)	0.05 (0.09)	0.005 (0.60)	0.14 (0.15)	0.11 (0.01)	0.001 (0.85)	0.47 (<0.001)	0.56 (<0.001)

Note: * = $p < 0.05$, ** = $p < 0.025$, *** = $p < 0.01$, one-tailed.
ITS: Interpersonal Trust Scale; GSS: General Social Survey.

Table 6. Mediation analysis using ordinary least squares regression for Conflict in Study 1; standard errors are given in brackets.

Dependent→	Model I	Model IIa	Model IIb	Model III	Model IVa	Model IVb	Model Va	Model Vb
	Specific Trust	Specific Trust	Specific Trust	Contribution	Contribution	Contribution	Contribution	Contribution
Constant	−4.30 (5.05)	−3.83 (5.68)	−4.07 (5.14)	2.42 (6.87)	−1.56 (7.63)	3.42 (6.91)	1.96 (5.64)	7.11 (5.19)
# Friends	0.57 (0.41)	0.57 (0.41)	0.53 (0.42)	0.54 (0.55)	0.50 (0.55)	0.39 (0.57)	−0.02 (0.41)	−0.09 (0.43)
Sex	0.07 (0.66)	0.09 (0.67)	0.05 (0.67)	0.58 (0.90)	0.43 (0.90)	0.46 (0.90)	0.34 (0.66)	0.42 (0.67)
Age	0.43* (0.24)	0.44* (0.24)	0.43* (0.24)	0.14 (0.32)	0.10 (0.32)	0.14 (0.32)	−0.30 (0.25)	−0.25 (0.25)
Harmony	−0.12 (0.66)	−0.11 (0.67)	−0.11 (0.66)	−0.79 (0.89)	−0.91 (0.90)	−0.72 (0.89)	−0.81 (0.66)	−0.63 (0.67)
First								
Experiment	−0.34 (0.78)	−0.35 (0.79)	−0.37 (0.79)	−0.62 (1.06)	−0.56 (1.06)	−0.73 (1.06)	−0.24 (0.78)	−0.40 (0.79)
First								
ITS		−0.01 (0.04)			0.07 (0.06)		0.08* (0.04)	
GSS Trust			−0.28 (0.80)					−0.96 (0.80)
Specific Trust							0.92*** (0.14)	0.91*** (0.14)
Δ R ² (p)	0.07 (0.57)	0.001 (0.85)	0.002 (0.73)	0.06 (0.65)	0.02 (0.24)	0.02 (0.26)	0.43 (<0.001)	0.42 (<0.001)

Note: * = $p < 0.05$, ** = $p < 0.025$, *** = $p < 0.01$, one-tailed.
ITS: Interpersonal Trust Scale; GSS: General Social Survey.

on the control variables Age, Sex (1 = male, 0 = female), Number of Friends, Experiment First and Harmony First. Number of Friends measures how many of the other participants the present the focal participant considers a friend. Experiment First is a dummy taking the value '1' when Task 1 came before Tasks 2 and 3, and '0' otherwise. Harmony First is a dummy taking the value '1' when Task 2 preceded Task 3, and '0' otherwise. In Model IIa and Model IIb, the ITS and the GSS trust question are added to the independent variables of Model I, respectively. Thus, Model IIa and Model IIb test the first step of the mediation of H5.

In Model III participants' contributions to the group public good are regressed on the control variables. In Model IVa the ITS is added as an independent variable to Model III. In Model IVb the GSS trust question takes the place of the ITS. Thus, in Models IVa and IVb the relation between General Trust and Contribution (H4) is investigated.

Finally, in Models Va and Vb Specific Trust is added as an independent variable to Models IVa and IVb, respectively. Thus, in these models it is investigated whether the effect of General Trust on Contribution indeed runs via Specific Trust (H5), and whether Specific Trust has a direct effect on Contribution (H2). Tables 5 and 6 show the results of these analyses for Harmony and Conflict, respectively.

H2 asserts that contribution to the group public good is positively related to the expectation that other participants contribute, and is corroborated by the data. Models Va and Vb of Tables 5 and 6 show large positive effects of Specific Trust on Contribution. Adding Specific Trust to Models IVa and IVb also yields large significant increases in *R*-square.

H4 argues that General Trust should be positively related to Contribution in Harmony, and negatively so in Conflict. Model IVa in Table 5 shows that the Harmony part of this hypothesis is corroborated for the ITS. For the GSS trust question (Model IVb) this is not the case. Models IVa and IVb in Table 6 show that the Conflict part of H4 is confirmed for neither the ITS nor the GSS trust question.

Testing of H4 showed that a direct effect of General Trust on Contribution exists for the ITS, in Harmony. Model IIa of Table 5 shows a positive effect of the ITS on Specific Trust, as required by H5. Finally, comparing Models IVa and Va of Table 5 reveals that the effect of the ITS on Contribution is mediated by Specific Trust (Sobel test statistic = 1.72, SE = 0.04, $p = 0.04$, one-tailed), corroborating H5 for this scale. The parameter for the ITS is reduced from 0.15 (in Model IVa) to 0.08 (in Model Va). Note that after adding Specific Trust in Model Va, a significant direct effect of the ITS remains.

Comparing Models IVa and Va of Table 6 we find an unexpected result. Although there is no direct effect of the ITS on Contribution in Conflict

Table 7. Correlations used for the Z-transformed Pearson–Filon test of H3 in Study 1.

	Contribution Harmony	Contribution Conflict	Specific Trust Harmony
Contribution Conflict	0.42		
Specific Trust Harmony	<u>0.79</u>	0.36	
Specific Trust Conflict	0.29	<u>0.67</u>	0.42

(contradicting H4), a significantly *positive* effect shows up when Specific Trust is added to the model. Thus, adding Specific Trust to the model disentangles the indirect and direct paths from the ITS to Contribution, rendering the latter path significantly positive. Thus, the Conflict parts of H4 and H5 must be rejected.

Finally, I test H3 using the Z-transformed Pearson–Filon (ZPF) test for comparing non-overlapping correlations (Raghunathan et al., 1996). Table 7 shows the correlation matrix for Contribution and Specific Trust in Harmony and Conflict. The difference between the two underscored correlations is significant, with ZPF = 1.61 ($p = 0.05$), corroborating H3.

Discussion of Study I

Summarizing the hypothesis testing, I find qualified support for H1. In addition, H2, H3 and the Harmony parts, but not the Conflict parts, of H4 and H5 are corroborated. The fact that H1 was corroborated is surprising from the point of view of the minimal-group paradigm, which would predict the opposite effect from the one found: higher contributions in Conflict than in Harmony. The corroboration of H2 shows that specific trust has a big impact on contributions, as was found in earlier research. The results pertaining to H3 suggest there is heterogeneity in terms of in-group bias, since the prediction that the correlation between specific trust and contributions will be higher in Harmony than in Conflict is derived from this assumption. Finally, H4 and H5 offer some first support for the claim that general trust positively affects contributions to public goods, an effect that is mediated by specific trust. Surprisingly, after adding specific trust to the model for Conflict, a positive effect of the ITS turned up, suggesting that scores on this scale are partly related to a tendency to contribute to in-group public goods regardless

of the effects on out-groups. This renders it questionable that high scores on this general trust scale indicate much social capital from a societal perspective.

Based on the results of Study 1, I conducted a second experimental study. There are two reasons for doing Study 2. The first is replication of the results of Study 1. The fact that H1 runs counter to the prediction from the well-established minimal-group paradigm in itself warrants replication. However, also the other hypotheses will be tested again. Related to this, as I explained in the *Methods* section, I used multiple general trust scales, of which I only retained the ITS and the GSS trust question for the models. The effects found for the ITS could thus be due to chance capitalization. The second reason to do Study 2 is that Study 1 did not involve measurement of in-group bias or 'group perspective'. Using a social preference instrument I especially designed for this purpose, I will do this in Study 2.

Study 2

Hypotheses. The distinction made in Table 3 between reciprocal players with a 'group perspective' and those with a 'society perspective' can theoretically be decomposed in two parts. The first part consists of an individual's 'general social preferences', that is, the extent to which an individual cares about the payoffs received by a general other. Positive general social preferences imply that an individual is willing to accept some (possibly very small) reduction of her own payoffs in return for an increase in the payoffs of someone else. Since a contribution to a public good is just such a transfer, individuals with positive general social preference can be expected to make such contributions.

Individuals holding a society perspective have these general social preferences regardless of whether the other is an in-group member or an out-group member. Individuals holding a group perspective, however, have an in-group bias. They value the payoffs received by fellow in-group members more than those received by out-group members.

In Harmony the effects of an individual's contribution to the in-group public good are positive on both the payoffs of in-group and out-group members. In Conflict, however, the situation is more ambiguous: in-group members are positively affected but out-group members negatively. This yields H6.

H6: The relation between general social preferences and contributions to the in-group public good is stronger positive in Harmony than in Conflict.

Logically, an individual's general social preferences and her expectations concerning the contributions made by others are unrelated. Much social-psychological research, however, suggests that empirically this is not the case. The 'false consensus effect' (Ross et al., 1977) contends that individuals expect others to hold the same social preferences as they themselves do. In the language of the previously mentioned social value orientation research, this means that individuals with a certain social value orientation will expect others to share this orientation (e.g. Bogaert et al., 2008; Kuhlman and Wimberley, 1976). Since the effects on the payoffs of others are 'unambiguously positive' (i.e. for both the in-group and the out-group) in Harmony, the false consensus effect would mean that general social preferences are positively related to the expectations concerning the contributions made by others. Combined with H6, this yields H7.

H7: The effect of general social preferences on contributions to the in-group public good in Harmony is mediated by the expectations concerning the contributions of others.

In Harmony, the effect of one's contribution on the payoffs of out-group members is positive. For individuals with a strong in-group bias this is less of an incentive to contribute than for individuals with little or no in-group bias. In Conflict, the effect of one's contributions on the payoffs of out-group members is negative. For individuals with a strong in-group bias this provides less of a *disincentive* to contribute than for individuals with weaker in-group bias. This yields H8.

H8: In-group bias is negatively related to contributions to the in-group public good in Harmony, and positively related to such contributions in Conflict.

Methods

In this section I describe how the methods employed in Study 2 deviated from those of Study 1. Therefore, I lump Design, Materials and Procedure together in one section.

Design, materials and procedure

The payoff parameters in Study 2 were identical to those of Study 1, and I again used a within-subjects design in which each participant completed four tasks. Task 1 was filling out the ITS scale, which had a Cronbach's

alpha of 0.69 in Study 2. Since they had hardly any effect in the final models of Study 1, and since I needed additional time for the social preference task (see Task 2), I did not measure any background variables such as sex and age. Task 2 consisted of three social preference measures. Each measure asked the participant to make seven choices between two allocation alternatives. Alternative 1 was the same throughout, and consisted of giving 8 points to some other participant and receiving 8 points oneself. Alternative 2 consisted of receiving X points oneself, with the other participant receiving zero points. In the seven allocations, X varied in integer numbers from 7 to 13, in the order 10, 8, 12, 7, 13, 9, 11. This social preference measure was not reciprocal: although each participant was 'the other' for someone else, participants were not reciprocally paired (i.e. no one was 'the other' of one's 'other'). Participants were ignorant of this matching and were only aware of the fact that they themselves had to make seven choices.

Participants completed Task 2a before they knew about the group structure. In Task 2a the other was a randomly drawn other participant from the room, and this task was intended to measure general social preferences. Tasks 2b and 2c were completed after the participants learned that there were two groups that would affect each other in terms of payoffs. It was stipulated that the size and sign of this effect would be explained later. Thus, participants knew the groups would affect each other, but not how. In Task 2b the other was an in-group member and in Task 2c the other was an out-group member. Tasks 3 and 4 were the same as Tasks 2 and 3 from Study 1, respectively (i.e. contributing in Harmony and Conflict, and guessing the contributions of the others).

One point scored in the experiment of Study 2 was worth 10 eurocents. Participants earned points in the same way as in Study 1. In addition they earned points in the social preference measures of Task 2. From each of the three sets of seven allocation questions (Tasks 2a, 2b, 2c) the computer, for each participant separately, randomly drew one question and paid out the participant and the other according to the choice made by the participant. Participants could earn up to a 39 additional points with this. The information and feedback conditions were identical to those in Study 1.

Eight sessions were organized in which 12 participants took part. The order of the tasks was randomized, with half the sessions having Task 1 at the beginning and half having Task 1 at the end, and half having Task 3 before Task 4 (Harmony first) and the other half having Task 4 before Task 3 (Conflict first). Task 2 always came before Tasks 3 and 4.

The procedures were identical to those of Study 1, with the additional on-screen explanation of the social preference measures in Task 2. Again, no participants had any trouble understanding the instructions. A session

Figure 3. Number of participants out of 96 that chose Alternative 2 (X for self, 0 for other) over Alternative 1 (8 for self, 8 for other) in the social preference measure of Study 2; the right-most bar represents Task 2a (general other), the middle bar represents Task 2b (in-group other) and the left-most bar represents Task 2c (out-group other).

lasted for about 35 minutes and participants earned an approximate average of 6.40 Euros.

Participants

There were 96 participants in the experiment. They were recruited from an on-line participant pool called the Sociological Laboratory (<http://www.gmw.rug.nl/~orsee/public/index.php>), and had not participated in the experiment of Study 1.

Results of Study 2

Description of preference types: in-group bias

The average observed contributions in Harmony and Conflict were 6.10 (SD = 3.50) and 5.61 (SD = 3.45), respectively. Of the 96 participants, there were only three truly selfish participants who contributed nothing in both Harmony and Conflict. Instead of looking at overinvestment and underinvestment as in Study 1, the data from Task 2 of Study 2 enable us to analyse in-group bias more directly.

Figure 3 shows the number of participants making the ‘individualist’ choice (i.e. Alternative 2) in each of the subtasks of Task 2. In Figure 3, I arranged the social preference questions in increasing order of the number

of points a participant could get when making the individualist choice. Thus, the first question in from the origin (question 4) had 7 for self and 0 for the other as the individualist choice, while the last question (question 5) had 13 for self and 0 for the other as the individualist choice.

As expected, the number of participants making the individualist choice increases, going from left to right (some participants individually did make some 'Gutman errors', though). Figure 3 shows that general social preferences, elicited with Task 2a, hardly differed from social preferences for in-group members (Task 2b). Indeed, the average number of times the individualist choice was made was identical (2.4) in these two tasks. However, in Task 2c participants substantively more often made the individualist choice, indicating the presence of in-group bias. Indeed, the average difference between the number of individualist choices in Tasks 2b and 2c is 1.09 ($SD = 1.53$), and is significant (paired sample t -test, $t_{95} = 7.01$, $p < 0.001$, two-tailed).

In the remainder I will use these data as follows. The number of times (out of seven) a participant chose Alternative 1 (i.e. the alternative that gives both the participant and the other 8 points) in Task 2a is taken to measure that participant's Social Preferences. The difference between this number in Tasks 2b and 2c is taken to measure the participant's degree of in-group bias.

Other descriptive results

To shed some first light on the main relations of interest in Study 2, Figures 4 and 5 show the scatter plots of participants' contributions in Harmony and Conflict related to their trust in others, their social preferences and their in-group bias.

The figures show that, similar to Study 1, in both Harmony and Conflict specific trust is strongly related to contributions, but general trust (the ITS) is not. The two specific trust measures (related to in-group and out-group members) are again strongly related in both conditions, but seem unrelated to the ITS.

Hypothesis testing

I start with replicating the hypothesis tests of Study 1 for the data of Study 2. The mean contribution of 6.10 ($SD = 3.50$) in Harmony is slightly higher than the mean contribution of 5.61 ($SD = 3.45$) in Conflict, as predicted by H1, and contradicting the alternative minimal-group hypothesis. However, this difference is not significant (paired sample t -test, $t_{95} = 1.15$, $p = 0.13$,

Figure 4. Scatterplot matrix main variables in Study 2, Harmony: ConHarmony: contributions in Harmony; ITS: Interpersonal Trust Scale; SpecTrustINH: specific trust in in-group members in Harmony; SpecTrustOUTH: specific trust in out-group members in Harmony.

one-tailed). As in Study 1, there is an order effect. When Task 3 occurred before Task 4 (i.e. Harmony before Conflict), the means were 6.25 (SD = 3.32) and 5.38 (SD = 3.73) for Harmony and Conflict, respectively, which is a marginally significant difference (paired sample *t*-test, $t_{47} = 1.41$, $p = 0.08$, one-tailed). When Task 4 occurred before Task 3 (i.e. Conflict before Harmony), the means were 5.96 (SD = 3.70) and 5.85 (SD = 3.17) for Harmony and Conflict, respectively, which is an insignificant difference (paired sample *t*-test, $t_{47} = 0.18$, $p = 0.43$, one-tailed). Thus, I again find qualified support for H1.

Since the means and variances of the contributions in both Harmony and Conflict were very similar in Study 1 and Study 2, I also pooled the data to test H1. In the pooled data, the mean contribution in Harmony is 6.10 (SD

Figure 5. Scatterplot matrix main variables in Study 2, Conflict: ConConflict: contributions in Conflict; ITS: Interpersonal Trust Scale; SpecTrustINC: specific trust in in-group members in Conflict; SpecTrustOUTC: specific trust in out-group members in Conflict.

= 3.47), and the mean contribution in Conflict is 5.51 (SD = 3.39), which is significantly different (paired sample t -test, $t_{154} = 1.83, p = 0.03$, one-tailed), reflecting the higher power due to the higher number of cases. The order effect is also found in the pooled data, with a significant difference in the predicted direction when Harmony was played first (mean difference = 1.23, paired sample t -test, $t_{77} = 2.71, p = 0.004$, one-tailed), and no difference when Conflict was played first (mean difference = -0.06, paired sample t -test, $t_{77} = -0.15, p = 0.88$, two-tailed). Finally, I test H1 by considering only the first game played, in the pooled data. The mean contributions in Harmony and Conflict are then 6.41 (SD = 3.31) and 5.85 (SD = 3.14),

which is, although in the predicted direction, an insignificant difference (independent samples *t*-test, $t_{154} = 1.09$, $p = 0.14$).

I test Hypotheses 2 and 5–8 by conducting mediation analyses using OLS regression, reported in Tables 8 and 9. As in Study 1, participants' estimates of the contributions by in-group members and out-group members were highly correlated (Pearson correlations of 0.94 and 0.92, $p < 0.001$ in Harmony and Conflict, respectively), as suggested by Figures 4 and 5. For the analyses of Tables 8 and 9, I therefore again constructed a single variable: Specific Trust. The means of Specific Trust in Harmony and Conflict were 5.59 (SD = 2.63) and 5.38 (SD = 2.53), respectively. The two mediation analyses (one for Harmony and one for Conflict) have a similar structure as those in Study 1.

H2, that contribution to the group public good is positively related to the expectation that other participants contribute, is corroborated by the data for both Harmony and Conflict. Models VI of Tables 8 and 9 show large positive effects of Specific Trust on Contribution, and adding Specific Trust yields large increases in *R*-square compared to the respective previous models.

Models III–VI show that the effects of General Trust, measured by the ITS, that were found in Study 1 are not replicated in Study 2. H4 therefore has to be rejected for both Harmony and Conflict. Models I of Tables 8 and 9, moreover, show that General Trust is not related to Specific Trust in Study 2, leading to the rejection of H5.

In order to test H3 for Study 2, Table 10 shows the correlation matrix for Contribution and Specific Trust in Harmony and Conflict. The difference between the two underscored correlations is not significant, with ZPF = 0.40 ($p = 0.34$), refuting H3. An alternative test of H3 can be conducted by looking at the interaction effect of Specific Trust and in-group bias in Conflict. According to H3, we should expect a positive parameter here. However, this effect was not found (results not shown in Table 9). Thus, in terms of replication of Study 1, the results of Study 2 are that only the results pertaining to the testing of H1 and H2 and replicated. Thus, contributions are a little higher in Harmony than in Conflict (especially when played in that order) and Specific Trust has a positive impact on contributions in both Harmony and Conflict.

With respect to the social preference and in-group bias hypotheses of Study 2, Models IV of Tables 8 and 9 show that H6 is corroborated by the data of Study 2. General social preferences are indeed positively related to contributions in Harmony, but not in Conflict. H7 is also corroborated by these data. Model I of Table 8 shows that general social preferences are positively related to the expectation that others contribute, and Model VI

Table 8. Ordinary least squares regression for Harmony in Study 2; standard errors are given in brackets.

Dependent →	Model I	Model II	Model III	Model IV	Model V	Model VI
	Specific Trust	Contribution	Contribution	Contribution	Contribution	Contribution
Constant	5.00*** (1.66)	6.50*** (0.62)	6.79*** (2.12)	5.42*** (2.19)	5.62*** (2.21)	0.46 (1.34)
Harmony First	0.09 (0.54)	0.29 (0.71)	0.30 (0.72)	0.40 (0.71)	0.30 (0.72)	0.10 (0.42)
Experiment First	0.88* (0.54)	1.08 (0.71)	1.07 (0.72)	1.27* (0.71)	1.31* (0.72)	0.40 (0.42)
ITS	-0.002 (0.03)		-0.01 (0.04)	-0.01 (0.04)	-0.01 (0.04)	-0.01 (0.02)
Social preference	0.24* (0.14)			0.39** (0.19)	0.40** (0.19)	0.15 (0.11)
In-group bias					-0.13 (0.21)	-0.29*** (0.12)
Specific Trust						1.08*** (0.08)
Δ R ² (p)	0.05 (0.33)	0.03 (0.29)	<0.001 (0.89)	0.04 (0.04)	0.004 (0.52)	0.62 (<0.001)

Note: * = $p < 0.05$, ** = $p < 0.025$, *** = $p < 0.01$, one-tailed.
ITS: Interpersonal Trust Scale.

Table 9. Ordinary least squares regression for Conflict in Study 2; standard errors are given in brackets.

Dependent →	Model I	Model II	Model III	Model IV	Model V	Model VI
	Specific Trust	Contribution	Contribution	Contribution	Contribution	Contribution
Constant	6.77*** (1.61)	6.32*** (0.61)	6.43*** (2.10)	6.21*** (2.21)	6.63*** (2.22)	-0.81 (1.46)
Harmony First	0.11 (0.52)	-0.48 (0.70)	-0.48 (0.71)	-0.46 (0.71)	-0.67 (0.73)	-0.74* (0.44)
Experiment First	0.54 (0.53)	0.94 (0.70)	0.93 (0.71)	0.97 (0.72)	1.05 (0.72)	0.45 (0.44)
ITS	-0.03 (0.03)		-0.002 (0.04)	-0.003 (0.04)	-0.003 (0.04)	0.03 (0.03)
Social preference	0.06 (0.14)			0.06 (0.19)	0.07 (0.19)	0.01 (0.12)
In-group bias					-0.28 (0.21)	-0.22* (0.13)
Specific Trust						1.08*** (0.09)
Δ R ² (p)	0.02 (0.70)	0.02 (0.33)	<0.001 (0.96)	0.001 (0.74)	0.02 (0.17)	0.61 (<0.001)

Note: * = $p < 0.05$, ** = $p < 0.025$, *** = $p < 0.01$, one-tailed.
ITS: Interpersonal Trust Scale.

Table 10. Correlations used for the Z-transformed Pearson–Filon test of H3 in Study 2.

	Contribution Harmony	Contribution Conflict	Specific Trust Harmony
Contribution Conflict	0.28		
Specific Trust Harmony	<u>0.81</u>	0.17	
Specific Trust Conflict	0.25	<u>0.80</u>	0.35

shows that the latter expectations indeed mediate the effect of social preferences on contributions. With a Sobel test statistic of 1.70 ($SE = 0.15$), the path from social preferences to contributions, running via Specific Trust, is significant ($p = 0.04$, one-tailed).

Finally H8 can be evaluated by looking at Models VI of Tables 8 and 9. Table 8 shows that the Harmony part of this hypothesis is corroborated, as in-group bias is indeed negatively related to contributions. Model VI of Table 9, however, shows that the Conflict part of H8 must be rejected. In fact, the opposite effect from the one predicted is found. Thus, unexpectedly, in-group bias is negatively related to contributions in Conflict, where a positive relation was predicted.

Discussion of Study 2

In terms of replicating the results of Study 1, the most remarkable result of Study 2 was that the relationships between General Trust, as measured by the ITS, and contributions to the in-group public goods were not found in Study 2. In addition, although there was a considerable degree of in-group bias and variance in terms of in-group bias and general social preferences, H3 (based on this heterogeneity) was not corroborated. The finding from Study 1, that contributions were higher in Harmony than in Conflict when the conditions occurred in that order, was replicated. This provides a second piece of evidence against the prediction derivable from the minimum-group paradigm. In addition, the positive relation between the expected contributions of others and a participant’s own contribution also replicated.

Participants’ social preferences and their degree of in-group bias had an effect on their contributions. General social preferences were positively related to contributions in Harmony and unrelated to contributions in

Conflict (H6), reflecting the negative effects of contributions on out-group members in the latter condition. As predicted (H7), general social preferences were positively related to Specific Trust in Harmony, and the effect of general social preferences on contribution in this condition was mediated by Specific Trust.

Finally, the effects of in-group bias (H8) are harder to make sense of. Although in-group bias was negatively related to contributions in Harmony as predicted, it was also negatively related to contributions in Conflict. Thus, in-group bias seems to measure a tendency not to contribute to the in-group public good, regardless of whether one is in Harmony or in Conflict. Note that in-group bias and general social preferences were uncorrelated (Pearson correlation of 0.02, $p = 0.82$, two-tailed).

General discussion

Trust is a key element of social capital. It is frequently argued that communities exhibiting high levels of trust are better able to produce public goods than communities low on trust. However, not all public goods that benefit the group producing it are propitious for society at large. Therefore, if trust is associated with the production of public goods, regardless of whether these public goods help or harm out-groups or the general society, trust cannot always be considered a good thing from the societal perspective.

In this paper I made a distinction between benign and malignant public goods, the former having positive externalities for out-group members and the latter having negative externalities for out-group members. Using a theoretical model of heterogeneous reciprocity preferences, I derived hypotheses concerning the difference between contributions to in-group public goods in Harmony and Conflict and concerning the effects of general and specific trust, social preferences and in-group bias on these contributions. I conducted two paid computer-mediated experiments to test these hypotheses.

In both studies I found that contributions in Harmony were slightly higher than contributions in Conflict. Also in both studies I found an order effect: when the Harmony condition was played before the Conflict condition the difference in mean contributions was (marginally) significant. When the order of the conditions was reversed, however, there was no significant difference in mean contributions.

One possible explanation for this order effect is that participants have a tendency to contribute less in the second condition than in the first. This tendency could then interact with the tendency to contribute more in Harmony than in Conflict to produce the results. A second explanation is

that participants viewed their fellow group members and the members of the other group differently, depending on the order of the conditions. Participants who were first in the Harmony condition might have come to regard members of the 'other group' as 'friends' or 'partners'. These participants might not want to harm these friends in the subsequent Conflict condition and thus lowered their contributions in Conflict. Conversely, participants who were first in the Conflict condition might have come to regard members of the other group as 'enemies'. These participants might not want to aid those enemies, and thus did not increase their contributions in Harmony. Note that this explanation assumes that the history of play *did* have an effect, even though there was no feedback about previous play or payoffs in between conditions.

The results from both studies pertaining to H1 run counter to theory and research in both the 'minimal-group paradigm' and the 'team game paradigm'. These theories would predict that in situations like Conflict, mean contributions to the group public good should be *higher* than in situations like Harmony. The groups in the experiments I conducted were only defined in terms of the payoff structure and therefore truly 'minimal'. This appears not to be enough to spark a competitive attitude towards the out-group.

I have designed the experiments such that besides the payoff structure, groups had no other features that could facilitate the identification of group members with their group. For instance, participants were unaware of the identity of their fellow group members or the identity of members of other groups. It was also impossible for them to infer the group composition from their payoffs: they received only global feedback on their earnings in the experimental tasks, and this only after all tasks had been completed; in addition to earning money via their contribution decisions, participants also earned money by guessing the contributions of other participants and it was not specified which part of the total amount they had earned in which task. Finally and importantly, all decisions were one-shot, such that no common history of play could develop.

Nonetheless, as indicated by the second explanation of the order effect above, the fact that H1 was not corroborated when Conflict was played before Harmony, might indicate that at least some social identification process induced by the 'history of play' (i.e. the order of conditions) was going on. These results raise an interesting follow-up question: will repeated play of the Conflict game facilitate group identification and spark intergroup competition such that repeated game contributions in Conflict significantly *do* exceed those in Harmony?

The results from Study 1 and Study 2 taken together suggest there is but a weak relation between general trust and the production of benign public

goods. In the Harmony condition of Study 1 scores on the ITS were positively and significantly related to participants' contributions to the in-group public good and this effect was mediated by participants' expectations of the contributions of others. In the case of Conflict in Study 1, however, the ITS had an unexpected significantly positive direct effect on the contributions. In Study 2, moreover, none of these effects replicated and there was no effect of the ITS on contributions to the in-group public good. Answers to the GSS trust question, employed in Study 1 as a second measure of general trust, were unrelated to contributions to the in-group public goods in either Harmony or Conflict.

The tentative conclusion on this issue is that one should be careful to advance the relationship between trust and contributions to public goods as a reason to regard trust as an element of social capital. In the two studies of this paper, I found effects of general trust on contributions in only one. Moreover, these results suggest that general trust might be associated with a *group's* capability of solving public good problems. From the societal viewpoint, however, the situation is more ambiguous, since in Study 1 there was a positive effect of general trust on contributions in Conflict. Thus, when society is characterized by the existence of many groups producing public goods with negative externalities for others, high general trust (as measured by the ITS) might not be very desirable.

The fact that in Study 1 answers to the GSS trust question were not associated with participants' contributions nor to their expectations of contributions of others, runs counter to results reported by Helliwell (2003). He shows that international differences in average scores on this question are good predictors of differences in the proportion of purposefully 'lost' wallets that are returned with their contents intact. In the case of returning a lost wallet, however, there is no immediate public good game involved for the participant. Returning the wallet is costly for the finder (especially in terms of the opportunity costs of keeping the content of the wallet), but the finder in this particular instance is not better off if all other participants return the wallets they found. Thus, I argue that returning a lost wallet should be interpreted as an act of altruism or perhaps trustworthiness more than as a contribution to a public good. Of course in the long run, when everybody has likely lost something of value sometime, all inhabitants might well be better off if everybody returned the things they found.

The expectations concerning the contributions of others, what I called 'specific trust', were strongly related to contributions to the in-group public goods in both Harmony and Conflict, and in both studies. Thus, specific trust that others contribute is a very good predictor of contributions to in-group public goods. However, this is true also in the case that the in-group

public good has large negative externalities for out-groups. This kind of specific social trust is therefore not really benign, and is not part of the social capital of the society at large.

General social preferences, as investigated in Study 2, are benign from a societal perspective, as they were positively related to contributions in Harmony and unrelated to contributions in Conflict. In addition, specific trust was a mediator of this effect in Harmony. Thus, although specific trust itself is not specifically benign, it is an important conduit for the relationship between social preferences and contributions to benign public goods.

Finally, in Study 2 I found that in-group bias was negatively related to contributions in both Harmony and Conflict. While the former effect was hypothesized, the latter is harder to explain. One design feature related to this effect might be that while the social preference and in-group bias measurements involved only *positive* payoffs in Task 2 of Study 2, contributing in Conflict implies inflicting *negative* payoffs on the out-group. Thus, even participants with a strong in-group bias might positively value payoffs for out-group members, making them reluctant to contribute to their in-group public good in Conflict. Note, however, that this does not explain why the *relationship* between in-group bias and contributions should be negative in Conflict. Measuring in-group bias with negative payoffs for the other might provide a first step to clarifying this issue. More generally, this paper has shown that to understand contributions to in-group public goods, we not only want to know to what extent a particular individual has reciprocal preferences (e.g. Fischbacher et al., 2001), but also to what extent this individual takes a group or society perspective in intergroup interactions.

Connected to the issue of heterogeneous reciprocity preferences and in-group bias is the effect of punishment on contributions in Harmony and Conflict. Fehr and Gächter (2002) have shown that in a regular public goods game the possibility of punishment provides reciprocal participants with an instrument to discipline selfish participants, without themselves decreasing their contributions to the public good. The effect of this is increased contributions by all group members. An interesting question for future research is now what the effects of punishment are in the Harmony and Conflict conditions of this paper. In particular, the case of Conflict seems very interesting, as one would expect reciprocal players that take a society perspective to punish players (both from their own group and from the other group) who *do* contribute. Reciprocal players who take a group perspective, however, should punish fellow group members who *do not* contribute, and possibly also punish members of the other group who *do* contribute. Thus, different player types could be distinguished by their punishment patterns.

The current paper has strong links with the existing literature on inter-group conflict, and it's instructive to discuss one of those in a little more detail. Halevy et al. (2008) presented an experimental design in which participants who decided to contribute to the collective good of their in-group had two options. They could either contribute in such a way that payoffs of out-group members decreased or in such a way that the out-group was unaffected. It turned out that participants rarely chose the option that harmed the out-group. Furthermore, contributions resembled those from classical intergroup conflict games, supporting the interpretation that these latter contributions were not based on the intention to decrease payoffs of out-group members, but by the desire to increase pay-offs of in-group members.

Notwithstanding the fact that this design is an excellent device to study the theoretical issue of purely negative out-group preferences, participants' behaviour in it is susceptible of social desirability. Making a contribution that harms the out-group, while another option that benefits the in-group equally but does not harm the out-group is available, is normatively very much disapproved of. Even spiteful individuals that positively hate the out-group might refrain from this course of action because of social desirability concerns. In the experimental design employed in the current studies, this is much less of an issue, since contributions to the group public good can easily be justified by the desire to help the in-group. Note, however, that the fact that contributions in Conflict were not higher than those in Harmony does chime with the results reported by Halevy and colleagues: the positive contributions in Conflict seem not to be driven by a desire to hurt the out-group (in which case they should be higher than the contributions in Harmony). It is by comparing results from different experimental setups that we eventually learn about individuals' motivations in situations like Conflict and Harmony.

Of course, the current studies are limited in a number of respects. One of the main ones is the fact that the trust scales that I used to measure general trust had low reliability. Even the ITS scale, that I eventually retained, did not do so well in this respect. A possible cause of this being the composition of the sample of participants (only undergraduate students), I intend to enlarge the participant pool from which the participants were drawn to include non-students as well.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Notes

1. I assume there is at least some actual heterogeneity with respect to reciprocity and perspective taking in the population. Not doing this would turn the strict inequalities of H1, H2 and H3 into equalities. It would not change the direction of the inequality.
2. Note that $a = -\frac{3}{2} < \frac{1-c}{G-1} = -1 < 0$, as required by the formal description of Conflict in the theory section.

References

- Allport GW (1954) *The Nature of Prejudice*. Cambridge, MA: Addison-Wesley.
- Anderson LR, Mellor JM and Milyo J (2004) Social capital and contributions in a public-goods experiment. *American Economic Review* 94(2): 373–376.
- Balliet D, Parks C and Joireman J (2009) Social value orientation and cooperation in social dilemmas: a meta-analysis. *Group Processes & Intergroup Relations* 12(4): 533–547.
- Bogaert S, Boone C and Declerck C (2008) Social value orientation and cooperation in social dilemmas: a review and conceptual model. *British Journal of Social Psychology* 47: 453–480.
- Bolton GE and Ockenfels A (2000) ERC: a theory of equity, reciprocity and competition. *American Economic Review* 90: 166–193.
- Bornstein G (1992) The “Free Rider” problem in intergroup conflicts over step-level and continuous public goods. *Journal of Personality and Social Psychology* 4: 597–606.
- Bornstein G (2003) Intergroup conflict: individual, group and collective interests. *Personality and Social Psychology Review* 7(2): 129–145.
- Bornstein G and Ben-Yossef M (1994) Cooperation in intergroup and single-group social dilemmas. *Journal of Experimental Social Psychology* 30: 52–67.
- Bornstein G, Kugler T, Budescu DV, et al. (2008) Repeated price competition between individuals and between teams. *Journal of Economic Behavior & Organization* 66: 808–821.
- Bouma J, Bulte E and van Soest D (2008) Trust and cooperation: social capital and community resource management. *Journal of Environmental Economics and Management* 56: 155–166.
- Brewer MB (1999) The psychology of prejudice: ingroup love or outgroup hate? *Journal of Social Issues* 55(3): 429–444.
- Brown R (1986) *Social Psychology*. 2nd ed. New York: Free Press.
- Camerer CF (2003) *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press/New York: Russell Sage.
- Choi JK and Bowles S (2007) The coevolution of parochial altruism and war. *Science* 318: 636–640.
- Coleman JS (1988) Social capital in the creation of human capital. *American Journal of Sociology* 94: S95–S120.
- Dawes RM (1980) Social dilemmas. *Annual Review of Psychology* 31: 169–193.

- Falk A and Fischbacher U (2006) A theory of reciprocity. *Games and Economic Behavior* 54: 293–315.
- Fehr A and Fischbacher U (2002) Why social preferences matter: the impact of non-selfish motives on competition, cooperation and incentives. *Economic Journal* 112: C1–C33.
- Fehr E and Gächter S (2002) Altruistic punishment in humans. *Nature* 415: 137–140.
- Fehr E and Gintis H (2007) Human motivation and social cooperation: experimental and analytical foundations. *Annual Review of Sociology* 33: 43–64.
- Fehr E and Schmidt KM (1999) A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114: 817–868.
- Fischbacher U (2007) Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10: 171–178.
- Fischbacher U, Gächter S and Fehr E (2001) Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters* 71: 397–404.
- Frank RH (1988) *Passions within Reason: The Strategic Role of the Emotions*. New York: W.W. Norton & Company.
- Fujiwara T and Kawachi I (2008) Social capital and health: a study of adult twins in the U.S. *American Journal of Preventive Medicine* 35(2): 139–144.
- Galea S, Karpati A and Kennedy B (2002) Social capital and violence in the United States, 1974–1993. *Social Science & Medicine* 55: 1373–1383.
- Glaeser EL, Laibson DI, Scheinkman JA, et al. (2000) Measuring trust. *Quarterly Journal of Economics* 115(3): 811–846.
- Halevy NG, Bornstein G and Sagiv L (2008) ‘In-Group Love’ and ‘Out-Group Hate’ as motives for individual participation in intergroup conflict: a new game paradigm. *Psychological Science* 19(4): 405–411.
- Hechter M, Friedman D and Kanazawa S (1992) The attainment of social order in heterogeneous societies. In: Coleman JS and Fararo TJ (eds) *Rational Choice Theory: Advocacy and Critique*. London: SAGE, pp. 79–97.
- Helliwell JF (2003) How’s life? Combining individual and national variables to explain subjective well-being. *Economic Modelling* 20: 331–360.
- Kawachi I, Kennedy BP, Lochner K, et al. (1997) Social capital, income inequality and mortality. *American Journal of Public Health* 87(9): 1491–1498.
- Kiyonari T, Tanida S and Yamagishi T (2000) Social exchange and reciprocity: confusion or a heuristic? *Evolution and Human Behavior* 21: 411–427.
- Knack S and Keefer P (1997) Does social capital have an economic payoff? A cross-country investigation. *Quarterly Journal of Economics* 112(4): 1251–1288.
- Kuhlman DM and Wimberley DL (1976) Expectations of choice behavior held by cooperators, competitors, and individualists across four classes of experimental game. *Journal of Personality and Social Psychology* 34(1): 69–81.
- Mulder L, van Dijk E, De Cremer D, et al. (2006) Undermining trust and cooperation: the paradox of sanctioning systems in social dilemmas. *Journal of Experimental Social Psychology* 42: 147–162.
- Olson M (1982) *The Rise and Decline of Nations*. New Haven, CT: Yale University Press.

- Putnam RD (1995) Bowling alone: America's declining social capital. *Journal of Democracy* 6(1): 65–78.
- Putnam RD (2007) *E Pluribus Unum: Diversity and Community in the Twenty-First Century. The 2006 Johan Skytte Prize Lecture*. *Scandinavian Political Studies* 30(2): 137–174.
- Rabin M (1993) Incorporating fairness into game theory and economics. *American Economic Review* 5: 1281–1302.
- Rabin M (2002) A perspective on psychology and economics. *European Economic Review* 46: 657–685.
- Raghuathan TE, Rosenthal R and Rubin DB (1996) Comparing correlated but non-overlapping correlations. *Psychological Methods* 1(2): 178–183.
- Rilling JK, Gutman DA, Zeh TR, et al. (2002) A neural basis for social cooperation. *Neuron* 35: 395–405.
- Robinson JP, Shaver PR and Wrightsman LS (1991) *Measures of Personality and Social Psychological Attitudes*. San Diego, CA, USA: Elsevier Academic Press.
- Rosenberg M (1957) *Occupations and Values*. Glencoe, IL: Free Press.
- Ross L, Greene D and House P (1977) The 'False Consensus Effect': an egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology* 13: 279–301.
- Rotter JB (1967) A new scale for the measurement of interpersonal trust. *Journal of Personality* 35: 651–665.
- Rotter JB (1971) Generalized experiences for interpersonal trust. *American Psychologist* 26: 443–452.
- Simpson B (2004) Social values, subjective transformations, and cooperation in social dilemmas. *Social Psychology Quarterly* 67(4): 385–395.
- Tajfel H and Turner J (1979) An integrative theory of intergroup conflict. In: Austin WG and Worchel S (eds) *The Social Psychology of Intergroup Relations*. Monterey, CA: Wadsworth, pp. 33–47.
- Van Lange PAM, Bekkers R, Schuyt TNM, et al. (2007) From games to giving: social value orientation predicts donations to noble causes. *Basic and Applied Social Psychology* 29(4): 375–384.
- Yamagishi T, Terai S, Kiyonari T, et al. (2007) The social exchange heuristic: managing errors in social exchange. *Rationality and Society* 19(3): 259–291.

Appendix I

Interpersonal Trust Scale, GSS trust question and civic norms scale used in the experiment

Interpersonal Trust Scale. All questions have the following five answer categories:

1. strongly agree;
2. mildly agree;

3. agree and disagree equally;
4. mildly disagree;
5. strongly disagree.

Participants are asked to rate their agreement on the following statements.

1. Hypocrisy is on the rise in our society.
2. In dealing with strangers one is better off to be cautious until they have provided evidence that they are trustworthy.
3. This country has a dark future unless we can attract better people into politics.
4. Fear and social disgrace or punishment rather than conscience prevents most people from breaking the law.
5. Using the honour system of *not* having a teacher present during exams would probably result in increased cheating.
6. Parents usually can be relied upon to keep their promises.
7. The United Nations will never be an effective force in keeping world peace.
8. The judiciary is a place where we can all get unbiased treatment.
9. Most people would be horrified if they knew how much news that the public hears is distorted.
10. It is safe to believe that in spite of what people say most people are primarily interested in their own welfare.
11. Even though we have reports in newspapers, radio and TV, it is hard to get objective accounts of public events.
12. The future seems very promising.
13. If we really knew what was going on in international politics, the public would have reason to be more frightened than they now seem to be.
14. Most elected officials are really sincere in their campaign promises.
15. Many major national sports contests are fixed in one way or another.
16. Most experts can be relied upon to tell the truth about the limits of their knowledge.
17. Most parents can be relied upon to carry out their threats of punishments.
18. Most people can be counted on to do what they say they will do.
19. In these competitive times one has to be alert or someone is likely to take advantage of you.
20. Most idealists are sincere and usually practise what they preach.

21. Most salesmen are honest in describing their products.
22. Most students in school would *not* cheat even if they were sure of getting away with it.
23. Most repairmen will not overcharge even if they think you are ignorant of their specialty.
24. A large share of accident claims filed against insurance companies are phony.
25. Most people answer public opinion polls honestly.

GSS trust question. Generally speaking, would you say that most people can be trusted or that you can't be too careful in life. (Most people can be trusted; Can't be too careful).

Appendix 2

Example handout instructions of harmony (Task 2)

This is how the value of your group's good depends on the contributions made by members of your group and members of the other group.

This is how the value of the other group's good depends on the contributions made by members of your group and members of the other group.