The Role of Reputations in the Evolution of Human Cooperation

Karolina Sylwester

PhD thesis

University of Newcastle

September 2010

ABSTRACT

In human societies, cooperation between strangers flourishes despite the risk of being exploited. Correct evaluation of others' cooperative intentions aids in selecting partners for profitable interactions. Assessment of intentions can be made by (a) considering individuals' reputations gained through observed interactions with others and third-party information (gossip) or (b) interpreting immediate cues such as facial expressions and body language. I empirically investigated the role reputations play in human economic decisions. More specifically, I addressed research questions such as (1) how and why people manage cooperative reputations, (2) what role reputations play in partner choice, (3) whether reputations can stabilize cooperation in groups, (4) whether people have a memory bias for specific reputations and (5) whether the ability to assess trustworthiness in faces relates to mind reading skills known as Theory of Mind (ToM).

Student participants were recruited for five experiments, all involving the use of economic games to a greater or lesser extent. Depending on the study, participants either played social dilemma games in groups under various experimental conditions or performed individual tasks e.g. recalled information previously presented in different contexts or assessed photographed faces with regard to their cooperativeness and completed ToM tasks.

The results provide evidence for the existence of reputation-based partner choice ('competitive altruism'). Participants strategically invested in reputations and reaped benefits from such investments in the form of profitable interactions with the most desired partners. By varying endowments I demonstrated that resource inequalities affect the way people invest in reputation with low-resource individuals behaving in a relatively more generous way than their high-resource counterparts. Moreover, I showed that cooperation in social dilemmas can be stabilized by introducing reputational incentives in the form of partner choice. My results also suggest that people have a memory bias for information about uncooperative acts which is independent of the cooperativeness of the environment they are exposed to. I found no relationship between the ability to identify cooperative intentions in faces and ToM skills.

In summary, by unravelling the mechanisms behind reputational cooperation my thesis sheds light on the reasons for extensive cooperation among strangers observed in humans.

Acknowledgements

I would like to wholeheartedly thank Dr Gilbert Roberts for his supervision of this thesis. I am grateful that he (1) trusted me three years ago and gave me a chance to work on such an interesting project, (2) provided me with competent support whenever I needed his advice but at the same time allowed me to develop my own ideas and (3) was always friendly, easy-going and diplomatic in his criticism which made our meetings fruitful and enjoyable. I would also like to thank my second supervisor, Dr Daniel Nettle whose comments greatly improved the design of the studies, data analysis and interpretation. I need to give credits to my two assessors Dr John Lazarus and Dr Quoc Vuong who supervised my progress over the three years and provided insightful remarks during our annual meetings. Next, I am grateful to the members of the Centre for Behaviour and Evolution group who patiently listened to my talks and constructively criticized me. In particular, I thank Dr Candy Rowe and Dr Tom Smulders for their encouragement and support with Darwin-related fun activities I co-organized. Finally, I am grateful to my postgraduate and post-doctoral colleagues who, when asked, eagerly dissolved my language and statistical doubts.

More specifically, I am indebted to Jonathan Sayers for his invaluable help in collecting the data for the study described in Chapter 4 and assisting with Qualtrics programming for the study presented in Chapter 7. I would also like to thank Jennie Abbott and Annabel Meredith who served as the two speakers in the recorded conversations in the study described in Chapter 6. I am grateful to Prof. Robin Dunbar for providing me with the new version of the social-cognitive ToM task used in Chapter 7, to Dr Daniel Nettle for recording the ToM stories, and to Dr Kai Alter for making his sound-proof lab available to us for recordings. I thank Dr Jan Verplaetse for enabling me to use his face stimuli in the experiment described in Chapter 7. Last but not least I am grateful to Dr Minna Lyons for her contribution to manuscript writing for Chapter 7.

I kindly acknowledge the Institute of Neuroscience, Newcastle University which provided funding for this project.

Table of Contents

1. Introduction 1

1.1. Inspiration and scope 1

1.2. Outline and individual contributions 3

1.3. Methodological issues 4

2. From indirect reciprocity to competitive altruism: a review of studies on reputation-based cooperation 9

2.1. Introduction 9

2.2. The uniqueness of human reputation-based cooperation 9

2.3. Indirect reciprocity and its different meanings 12

2.4. Conceptual extensions of indirect reciprocity 14

2.5. Empirical evidence for indirect reciprocity 17

2.6. Theories behind reputation-based partner choice 21

2.7. Competitive altruism: theoretical and empirical evidence 24

2.8. Conclusions 26

3. Profitable cooperation through competitive altruism 28

3.1. Introduction 29

3.2. Study 1 30

3.2.1. Method 30

3.2.2. Results 33

3.3. Study 2 34

3.3.1. Method 34

3.3.2. Results 35

3.4. General discussion 37

4. The effect of resource inequality on reputation-based cooperation in social dilemmas 40

4.1. Introduction 41

4.1.1. Cooperation and fairness in social dilemmas 41

4.1.2. The role of reputation in social dilemmas and partner choice 42

4.1.3. Perceptions of pro-social attributions and intelligence 43
4.1.4. Outline and hypotheses 44

4.2. Method 45

4.2.1. Participants and payments 45
4.2.2. Procedure 45
4.2.3. Design and data analysis 48

4.3. Results 49

4.3.1. Cooperation cost and endowment 49
4.3.2. Partner preference 50
4.3.3. Trait assessment 51

4.4. Discussion 52

5. The effectiveness of competitive altruism and indirect reciprocity at sustaining cooperation in social dilemmas 55

5.1. Introduction 56
5.2. Method 58
5.3. Results 60

5.3.1. Main analysis 60
5.3.2. Additional analysis 65

5.4. Discussion 66

6. Memory bias for reputational gossip: Cheater-detection or rarity-detection? 69

6.1. Introduction 70
6.2. Methods 74

6.2.1. Participants 74
6.2.2. Material 74
6.2.3. Design and procedure 75

6.3. Results 76
6.4. Discussion 79

7. The role of Theory of Mind in assessing trustworthiness 83

7.1. Introduction 84

7.2. Method 87

7.2.1. Participants 87

7.2.2. Materials 87

7.3. Results...89

7.4. Discussion..91

8. Concluding remarks 95

8.1. Summary of results and implications...95

8.2. Limitations and future directions 98

Appendix A... 101
Appendix B 102
Appendix C... 104
Appendix D 105
REFERENCES 106

1. Introduction

1.1. Inspiration and scope

This thesis is a result of research I conducted between October 2007 and September 2010. When I started my PhD I was familiar with the traditional theories of cooperation such as kin selection (Hamilton, 1964) and reciprocal altruism (Trivers, 1971) but did not know much about studies on reputation-based cooperation. I had been, however, already attracted to the concepts of costly signalling (Smith and Bliege Bird, 2005) and the handicap principle (Zahavi and Zahavi, 1997) and intrigued by phenomena which, at least at first sight, cannot be explained by natural selection (Darwin, 1859). Research on reputation building behaviour appeared to be fascinating because it related to a general philosophy of perceiving human behaviour through an ecological and evolutionary lens. Such an approach assumes that all actions serve a purpose of efficiently and successfully dealing with the surrounding environment.

From my undergraduate work on displays of status, I could appreciate how much people can sacrifice in order to maintain high reputation. At that time, I had a chance to read into sociological views on human behaviour which highlight its plasticity in response to the presence and actions of others. The idea of building cooperative reputations ties in well with a broader concept of people actively directing their behaviour in front of an audience. In his famous sociological work, Goffman (1959) drew a parallel in which he compared life to a theatre and human behaviour to acting in a play in which actors take pains in order to dazzle the audience. Striking differences in how people behave under conditions of anonymity and when they are observed by others indicate that human behaviour is indeed shaped by perceptions of being monitored.

What kinds of benefits can people receive when acting in a certain way in the presence of others? Competitive altruism theory, which is the main topic of my thesis, provides a plausible answer to this question: in a world where collaborators can achieve much more than single individuals, acquiring a trustworthy and cooperative social partner considerably increases profits. In order to have access to such a partner an individual needs to prove that he himself will make a good collaborator. When behaviour is driven by competition for partners individuals try to outperform each other in terms of qualities valued in coalitions such as cooperation, fairness and generosity. The idea of costly altruistic displays corresponds to the concept of conspicuous consumption formulated by Veblen (1899). According to him, for maintaining high social status it does not suffice to merely accumulate goods, one also needs to spend them in the most ostentatious way. Similarly, an individual whose reputation is continuously assessed by others has to keep investing in it. In his book, Veblen made a loose link to conspicuous giving:

> *The aid of friends and competitors is therefore brought in by resorting to the giving of valuable presents and expensive feasts and entertainments. (…) Costly entertainments, such as the potlatch or the ball, are peculiarly adapted to serve this end. The competitor with whom the entertainer wishes to institute a comparison is, by this method, made to serve as a means to the end.* (p.47)

During my PhD I discovered the potential of reputations in explaining human cooperation. With extensive ways of monitoring others via language, people can rarely feel completely unobserved and free of reputational concerns, which affects how they behave. Obviously, kin selection and direct reciprocity still play an important role in human interactions, however, the willingness to trust strangers, even when high stakes are considered (e.g. when making online purchases), indicates the extent to which people rely on reputations. The original aim of this project was to empirically test the competitive altruism theory (CA), an idea conceptualized by my supervisor, Dr Gilbert Roberts (1998). I conducted three experiments examining the mechanism of CA under various conditions. In the course of my PhD I became interested in the ongoing debate on whether people are equipped with some special cognitive mechanisms that allow them to efficiently deal with social partners such as the cheater detection module (Cosmides and Tooby, 1992). At the same time, I was curious how reputational gossip is perceived and spread. I combined these two topics in an experiment examining memory biases in processing gossip. Another area that intrigued me was how people make judgements of someone's cooperative intentions when no reputational information is available. Recent reports indicate human proficiency in reading facial cues of cooperation and defection. However, if there was a pressure to evolve the ability to read cooperative intentions, another pressure should have promoted concealing such intentions, in particular the intention to cheat. People vary in the extent to which they can accurately predict others' trustworthiness. I investigated whether the variation in trustworthiness assessment might be related to the Theory of Mind (ToM) skills. Overall, this thesis provides new insights into how people manage their reputations and how concerns about reputations affect cooperation. Furthermore, it contributes to the knowledge about reputational gossip and discusses the link between cooperative intention recognition and ToM.

1.2. Outline and individual contributions

The chapters in this thesis were originally prepared as research articles. The two studies described in Chapter 3 have been published as Sylwester and Roberts(2010). Other studies have been written up as manuscripts and are currently under revision or in preparation for submission. Most of the studies have been presented as talks or posters at European and international conferences on human behaviour. The chapters are sometimes identical or very similar to the research manuscripts submitted. For all chapters I produced the first draft which was then revised by Dr Gilberts Roberts who is also the co-author of all the papers submitted for publication. For this reason in all studies a plural 'we' form is used. For some experiments (in particular Chapter 7) I also sought advice from my second supervisor Dr Daniel Nettle. Over the course of my PhD my work was monitored by the assessment panel, Dr John Lazarus and Dr Quoc Vuong, from whom I received detailed feedback. For the majority of the experiments I collected the data myself with the exception of Chapter 4 and Chapter 7. Jonathan Sayers volunteered to assist me with conducting the experiment described in Chapter 4. Under my supervision, he recruited participants, arranged experimental sessions and implemented the program written by me in Z-Tree software. Jonathan also contributed to programming the questionnaire in Qualtrics for the study described in Chapter 7. Chapter 6 describes a study on the relationship between ToM and the ability to assess trustworthiness. At the EHBEA conference in Wroclaw (2010) I discovered that Dr Minna Lyons from Liverpool Hope University had investigated a similar topic using different methodology. Because of the lack of an effect found we decided it may be beneficial if we combine our studies and publish them in one paper. Chapter 7 contains only the data collected by me, as I did not contribute in any way to experimental design and data collection of Dr Lyons' study. Chapter 7 is a modified manuscript that we intend to submit as a multiple study with Dr Lyons. I was the main person working on the manuscript, however Dr Lyons contributed to it and I received thorough feedback from Dr Daniel Nettle.

The original project I was appointed to work on during my PhD involved empirically examining CA, a theory of reputation-based cooperation. The literature review (Chapter 2) and the first three experiments (Chapters 3, 4 and 5) all relate to CA. Because they were written as separate papers there may be some overlap or repetition in terms of the presented background, which is however necessary in order to view the studies in the right context. As mentioned earlier, over the course of my PhD I became interested in whether people are equipped with some cognitive mechanisms that allow them to make good use of reputational information and also how people assess cooperativeness using some proximate cues such as facial expressions when no reputation or experience is available. I therefore conducted two additional studies (Chapter 6 and 7) which are not linked to CA but do provide answers to related research questions concerning cooperation. The common feature of the experiments included in this thesis is that to a greater or lesser extent they all try to answer a question of how people assess someone's future cooperativeness. Depending on the situation, people can use information about other's pro-sociality which they observed, recall reputational gossip about them or try to predict their behaviour from cues of trustworthiness.

This thesis is organised in the following way. In Chapter 2, I present a review of studies on reputation-based cooperation which contrasts two theories: indirect reciprocity (IR) and CA. A number of reviews of the traditional theories of cooperation exist in the literature (e.g. West et al., 2007), but there is no systematic review covering studies on cooperative reputations. The need for such a review arises also from the fact that IR and CA share some assumptions and therefore might be difficult to distinguish from each other. My review disentangles the two concepts and critically assesses the hitherto conducted research. In the next three chapters I present experiments investigating the mechanisms underlying reputation-based cooperation. In Chapter 3, I show how various reputational incentives affect cooperative behaviour and that investing in reputation translates to benefits in the form of access to cooperative partners and profitable interactions with them. In Chapter 4, I examine reputation-based cooperation in a situation where individuals vary in the amount of resources that they possess. I also investigate the role of the absolute and relative cost of reputational investments in partner choice. In Chapter 5, I compare the efficiency of the two mechanisms of reputation-based cooperation - CA and IR - at enhancing cooperation in social dilemmas. In Chapter 6 I discuss and empirically test the possibility of a memory bias for reputational gossip. Chapter 7 describes an investigation into the role of ToM in the ability to make accurate judgements about someone's cooperative intentions. Finally, in Chapter 8 I summarise my findings, consider the limitations of the studies and present ideas for future research.

1.3. Methodological issues

The studies presented in this thesis derive from different disciplines such as behavioural economics, social psychology, evolutionary psychology and experimental psychology. Experimental practices in psychology are more relaxed than those in economics in at least four respects: the use of scripts, monetary incentives, repetition and deception (Hertwig and Ortmann, 2001).

Economists commonly use scripts (detailed instructions) in which they clearly specify participants' roles in the game, their possible decisions and payoffs associated with different decisions. Psychologists, on the other hand, usually provide very general information about the study and allow participants to conjecture the context of the experiment. Hertwig and Ortmann (2001) assert that scripts increase replicability whereas the lack of them may lead to a change in participants' behaviour. Participants might be trying to guess the context of a task or perform in order to satisfy experimenter's expectations. While I agree that using scripts can reduce any noise resulting from different interpretations of the study, it is worth pointing out that some psychological experiments aim to measure participants' spontaneous judgements not embedded in any context (e.g. studies on attractiveness). Moreover, in real life people often act under conditions of uncertainty or in an unpredictable environment, hence measuring people's behaviour in vaguely determined contexts may faithfully reflect how people make real-life decisions. Also, researchers providing participants with as little information as possible avoid the problem of the framing effect. Different wordings used in the instruction can induce cooperation or competition and in consequence affect participants' behaviour (van Vugt, 2001). In this thesis, for the three experiments involving playing economic games with others (Chapters 3, 4 and 5), I adopted the rigorous economic conventions. Participants were provided with scripts and knew exactly in what way their payoffs would be affected by their decisions. For the studies described in Chapters 6 and 7 I used a traditional psychological methodology and provided participants with general instructions which made no reference to their role and context of the tasks.

Second, economic and psychological studies differ with regard to using repeated trials versus snapshot studies (Hertwig and Ortmann, 2001). Economists posit that participants may behave differently when exposed to a novel task for the first time than after habituating to it e.g. a common finding is that participants' contributions to a public pool decrease over time in iterated games (e.g. Isaac et al., 1994). In my opinion, choosing the right method depends on what a researcher aims to measure. In order to capture optimal behaviour repeated trials affected by learning should be examined whereas when spontaneous and intuitive behaviour is of interest snapshot studies would be appropriate. Henrich (2001) convincingly argues that in real life people often make one-shot decisions, hence experiments reflecting such decisions are justified. Henrich's et al. (2004) seminal cross-cultural work on human pro-sociality relied mostly on one-shot trials of the Ultimatum game because the researchers intended to capture natural, non-learned behaviour of inexperienced individuals. In this thesis, in the three experiments employing behavioural economics procedures (Chapters 3, 4 and 5) rounds were repeated and different conditions were presented in a balanced order. The possible memory effects resulting from such repetition are discussed in the greatest detail in Chapter 5.

The third difference between economic and psychological approaches to methodology is the use of financial incentives (Hertwig and Ortmann, 2001). Experimental economists typically use monetary incentives paid according to participants' performance (Camerer, 2003) while psychologists tend to provide participants with hypothetical rather than actual choices (Hertwig and Ortmann, 2001). There is mixed evidence with regard to how monetary incentives affect performance (Hertwig and Ortmann, 2001). One of the debated problems is the occurrence of a hypothetical bias (also referred to as 'cheap talk') – being more willing to declare to pay in a survey than when having an option of an actual payment. When considering contributions in a public goods game used frequently in research on human cooperation no strong evidence of a hypothetical bias exists. In a study by Champ and Bishop (2001) participants were asked whether they would be willing to purchase a public good, wind-generated electricity that would benefit the whole community. Half of participants were asked a hypothetical question whereas the other half had to actually pay the agreed amount if they decided for the purchase. Approximately a half fewer participants decided to invest in the public good when it involved actual expenses in comparison to when it was hypothetical. In contrast, a recent study involving participants playing a public goods game with heterogeneous resources did not show any evidence of such a hypothetical bias (Mitani and Flores, 2009). In the three experiments described in this thesis which employed the public good game, in most cases, participants received monetary payment in proportion to their experimental earnings (but see the detailed descriptions in methods of each chapter). Two anonymous reviewers who read two different manuscripts included in this thesis complained that the amounts given to participants were relatively low to induce 'true' economic behaviour (one of the reviewers was concerned that we did not examine the behaviour of *Homo economicus* but *Homo ludens*). Such an allegation does not appear to be valid. First, undergraduate students usually rely on their parents' financial support or undertake various part-time jobs; hence the amounts gained in an experiment are likely to have a higher value to them than to academic staff. Also, although average earnings reached £6 - £7, the maximum possible payoff was much greater. Therefore, although not very probable it *was* possible to earn substantial amounts of money, depending on others' behaviour (i.e. if the target individual was selfish and all others cooperated in each round). Second, economic experiments conducted in developing countries in which the experimental stakes were even equal to a monthly salary produced similar results to those with relatively low stakes (Camerer, 2003).

Finally, another methodological issue that differentiates psychologists from economists according to Hertwig and Ortmann (2001) is the use of deception. Deceiving participants can be questioned on internal and external grounds. Ethically speaking, deception may cause some discomfort, resentment and the impression of being 'fooled' in participants, as arrestingly illustrated by Milgram's (1963) famous study on obedience. However, a review of studies using deception did not allow for an unanimous conclusion of how participants' mood is affected (Hertwig and Ortmann, 2008). An external reason for not adopting deception is the possibility that it will increase participants' suspicion and distrustfulness which will affect the results of future studies (Hertwig and Ortmann, 2001). I recruited participants separately for each experiment presented in this thesis. For the experiments described in Chapters 3, 4 and 5, I ensured that participants were naïve i.e. took part only in one of the described experiments. Because Chapters 6 and 7 were conceptually and methodologically different from earlier chapters focusing on CA, some participants who did one of the first four experiments (two experiments from Chapter 3, one experiment in Chapter 4 and one in chapter 5) were able to take part in experiments in Chapters 6 and 7. In defence of deception it can be argued that, besides being often the least expensive method, without it, some important findings would not have been brought to light (see van Vugt, 2001). Following the APA recommendations I used experimental manipulation in a cost-effective way; that is, only in cases when it was absolutely necessary in order to induce conditions which are difficult to observe naturally (see Chapter 4). Participants were thoroughly debriefed after the experiment and did not express any complaints when the nature of the experiments was revealed to them. All presented experiments received ethics approvals from the Newcastle University Psychology Ethics Committee.

Recently, it has been pointed out that the majority of findings in behavioural science are obtained from a very non-representative group of Western, Educated, Industrialized, Rich, and Democratic (WEIRD) participants (Henrich et al., 2010). WEIRD participants can behave strikingly different from groups less well represented in studies (e.g. with small-scale societies) in domains including visual perception, spatial reasoning, sense of fairness and cooperation. The evidence presented by Henrich et al. (2010) is compelling but I believe studies using students allow one to make a first step when investigating a novel problem. As suggested by Gächter (2010), student participant pools make an excellent benchmark when examining important research questions in behavioural economics. Although the results presented in this thesis are based on data mostly from students and therefore cannot be generalized, they provide a useful first test of the proposed hypotheses.

2. From indirect reciprocity to competitive altruism: a review of studies on reputation-based cooperation

2.1. Introduction

Over the last three decades researchers have been accumulating evidence for a new explanation of the evolutionary puzzle of human cooperation – reputation building. The dawn of the investigation into reputation-based cooperation came with the publication of Richard Alexander's "The Biology of Moral Systems" (1987). Alexander's role in adding reputation building to an array of evolutionary explanations of cooperation relied on the fact that he analysed the possible costs and benefits of high and low reputations and discussed long-term consequences of both. The notion of a cooperative reputation refers to a characteristic of an individual which describes their willingness to cooperate. This characteristic is constructed by an exchange of information about an individual's past behaviour. Technically, a cooperative reputation consists of a history of cooperative and uncooperative acts towards other individuals. In practice, reputations are used to make predictions about the probability of future cooperative behaviour. Acquiring a high cooperative reputation is costly: an individual needs to invest in unreciprocated help towards others. What are the benefits of a high cooperative reputation? Two evolutionary theories discussed in this review, indirect reciprocity (IR) and competitive altruism (CA), propose different mechanisms to answer this question. The lack of cross-referencing does not allow the two theories to be put into perspective and necessitates a critical assessment summarising the contributions of IR and CA to understanding the role of reputations in cooperation.

2.2. The uniqueness of human reputation-based cooperation

In 'The Descent of Man, and Selection in Relation to Sex' Darwin (1871) stated: *"I fully subscribe to the judgement of those writers who maintain that of all the differences between man and the lower animals the moral sense or conscience is by far the most important"* (p.67). Although morality strictly refers to *"a code of conduct put forward by a society or (…) rational persons"* (Gert, 2008) it can more simply be viewed as a display of other-regarding preferences. Darwin (1871) clearly thought that morality contributes to human uniqueness in the animal kingdom but he also asserted that "*the difference in mind between man and the higher animals (…) is certainly one of degree and not of kind*" (p.101). Indeed, the idea that other-regarding preferences are confined to humans has been challenged by numerous studies highlighting various forms of empathy in animals (for a review see: de Waal, 2003). The presence of pro-social preferences in animals undermines the notion of the distinctiveness of human moral behaviour and encourages seeking the roots of morality in biology rather than culture. On the other hand, a study by Silk et al. (2005) highlighted chimpanzees' selfishness and the lack of concerns for others even when help is not costly.

Although cooperation appears to be difficult to reconcile with natural selection theory, some examples of cooperative behaviour have now clear evolutionary explanations. Undisputedly, kin selection theory or inclusive fitness: helping those who share the same genes, unravels the mechanism behind cooperation between related individuals e.g. consensual slavery in social insects or nepotism in humans (Hamilton, 1964). Similarly, reciprocal altruism or direct reciprocity, that is, exchanging cooperative acts between two individuals accounts for numerous examples of cooperation such as food sharing or warning calls (Trivers, 1971). However, there are limitations to the application of these theories: according to kin selection, a cooperative act should occur only when the cost to the benefactor is smaller than the benefit to the receiver multiplied by the coefficient of relatedness, whereas in reciprocal altruism the benefit to the receiver must exceed the cost of performing the act and the two individuals need to re-meet. Helping kin or reciprocating help can surely be called moral in contrast to not doing so but human morality extends beyond cooperation with immediate benefits. Kin selection and reciprocal altruism explain many non-human acts of cooperation or other-regarding preferences but do not allow for interpreting the 'stranger-regarding' behaviour which entails helping unrelated individuals who will not have a chance to reciprocate (Palameta and Brown, 1999).

The biological approach to cooperation and altruism relies on finding ways in which helpful individuals can profit from the help that they provide. Hence, the term 'altruism', "*helping … purely out of the desire to benefit someone else, with no benefit (and often a cost) to oneself*" (Aronson et al., 2004, p.382), may not be the most appropriate one to use when discussing evolutionarily based cooperation. Psychological altruism clearly exists in humans: people do not meticulously calculate whether an altruistic act will pay off or not, and their pro-social actions are proximately motivated by the feeling of 'warm glow' which results from the activation of reward circuits in the brain (Rilling et al., 2002). When considering biological altruism, however, an individual's fitness is treated as a reference point, hence a truly altruistic act would require that cooperation decreases the fitness of the benefactor and increases the fitness of the recipient (Wilson, 1975). Although psychological motivations for helping may not be dictated by future rewards, if the benefactor is rewarded and the cost of a cooperative act is recouped, the act cannot be called altruistic in the biological sense. For this reason, the names of biological theories which explain cooperation through payoffs to the benefactor such as 'reciprocal altruism' or 'competitive altruism' can be misleading and the use of the term 'cooperation' or 'aid' rather than 'altruism' would be more appropriate. At the same time, as noticed by de Waal (2008), cooperation can be truly other-oriented in the sense that it yields rewards via others.

When describing reciprocal altruism Trivers (1971) mentioned multiparty interactions relying on information exchange such as detecting cheaters and passing on their reputations. A clear theoretical framework of reputation-based cooperation was presented by Alexander (1987) who suggested that IR may be unique to humans; however, he did not exclude the possibility that animals may possess some rudimentary form of it. Indeed, a few reports indicate that animals use reputations when making decisions who to choose as a partner. Cleaner fish and reef fish form mutually beneficial relationships in which the cleaner fish feed by removing parasites and dead tissue from the reef fish (Bshary, 2002). Cleaners can exploit clients by feeding on their healthy instead of parasitized tissue which results in the clients jolting. Bshary (2002) observed that clients pay attention to the interactions between cleaner fish and other reef fish and make a decision who to invite as a cleaner based on cleaners' reputations. Later it was experimentally shown that cleaners observed to be cooperative were preferred by clients and that a cleaner's cooperation towards the current client increased when they were watched by other clients (Bshary and Grutter, 2006). The existence of reputation-based choices of partners has also been tested in great apes. Among gorillas, orang-utans, bonobos and chimpanzees only the last species was shown to spend significantly more time with and beg for food from a person observed to be cooperative than with a selfish person (Russell et al., 2008, Subiaul et al., 2008). The presented evidence from animals is scarce and does not directly address the concept of IR according to which individuals of high cooperative reputation would be more likely to receive help from others, or CA which posits that individuals try to outcompete each other with cooperative reputations in order to secure access to desired partners. On the contrary, the fish and ape studies mentioned above demonstrate that individuals of high reputation are more likely to be asked for help again. Hence, the current evidence does not allow for extending the theories of IR and CA to animals.

2.3. Indirect reciprocity and its different meanings

Give, and it shall be given unto you… For with the same measure that ye mete withal it shall be measured to you again. Luke 6:38

The idea that good deeds will be rewarded by a third party is not new. Abrahamic religions use this notion in order to encourage followers to cooperate: good deeds are believed to be reciprocated by the supernatural force in the afterlife. It was not until 1987, however, that Richard Alexander noticed that cooperation can be rewarded *hic et nunc* by someone other than the beneficiary. Interestingly, Alexander (1987) posited that the driving force for moral systems based on IR is between-group competition. Such an approach has probably been derived from Darwin (1871) who noted:

> *There can be no doubt that a tribe including many members who (…) were always ready to give aid to each other and to sacrifice themselves for the common good, would be victorious over most other tribes; and this would be natural selection.* (p.159)

Higher cooperation among in-group than out-group members is well documented in humans and occurs even when groups are created by using such an insignificant criterion as the preference of one painter over another one (Tajfel, 1970, Tajfel et al., 1971). Recently Yamagishi and Mifune (2008) showed that reputation-based cooperation is sensitive to the observer's group membership. In their study, in-group favourism disappeared when the recipient of the cooperative act was oblivious of the group membership of the benefactor. Only when the donor's membership was public did he invest in in-group reputation by giving more to in-group rather than out-group members. Similarly, when recipients of cooperation did not know the group identity of the donor, donors exhibited an in-group bias only when exposed to an image of eyes, a cue of being watched (Mifune et al., 2010). Such results indicate that individuals adjust their cooperativeness according to whether the observers belong to their group or not because higher in-group favourism is likely to pay off through IR within one's own social group.

Alexander (1987) described IR as cooperative behaviour where "*the return is expected from someone other than the recipient of the beneficence*" (p.85). Earlier in his book he provided two examples of IR: either "*A helps B, B helps C, C helps A*" or *"A helps B; C, observing, later helps A; A helps C"* (p.81). Reputation matters only in the latter example where individuals' behaviour is monitored and the decision to help is inspired by the high cooperative reputation of the recipient. It is worth noting that Alexander actually suggests that IR leads to direct reciprocity: after the initial assessment individuals C and A exchange help. In the former example cooperation is motivated by the mere fact of receiving help and is directed to an individual of an unknown reputation.

The two examples presented by Alexander have been dubbed *upstream* and *downstream* IR where upstream reciprocity refers to a unidirectional stream of cooperation whereas downstream reciprocity occurs when cooperation is initiated by acquiring positive reputational information about others (see Figure 2.1). Simulations of the evolution of downstream and upstream strategies demonstrated that with increasing group size, it becomes particularly difficult for upstream IR to dominate (Boyd and Richerson, 1989). Although downstream IR requires that individuals know how others behaved (rather than what happened to them), such a strategy can evolve under a wider range of conditions than downstream reciprocity. Upstream reciprocity has been referred to as 'generalized reciprocity' which clearly separates it from the reputation-based IR (Pfeiffer et al., 2005). The reason for the occurrence of generalized reciprocity is difficult to explain: why, after receiving help from a stranger, would an individual feel an 'urge' to help someone else? Nowak and Roch (2007) assert that generalized reciprocity is a by-product of direct reciprocity, a misdirected act of gratitude. A mechanism of generalized reciprocity does not require complex cognitive skills; individuals simply follow a rule: help if you have been helped. It has been shown that generalized reciprocity can induce and maintain cooperation in small groups (Pfeiffer et al., 2005), however later it was found that in order for cooperation to evolve, generalized reciprocity has to be linked to another mechanism (Nowak and Roch, 2007). Generalized reciprocity, unlike downstream IR, due to its simplicity, is a more plausible mechanism to be found in animals. Rutte and Taborsky (2007) demonstrated that rats were willing to help unknown individuals if they had been helped before. This review focuses on cooperation invoked by reputational concerns and not direct experience hence the term 'indirect reciprocity' (IR) is reserved to refer to downstream indirect reciprocity.

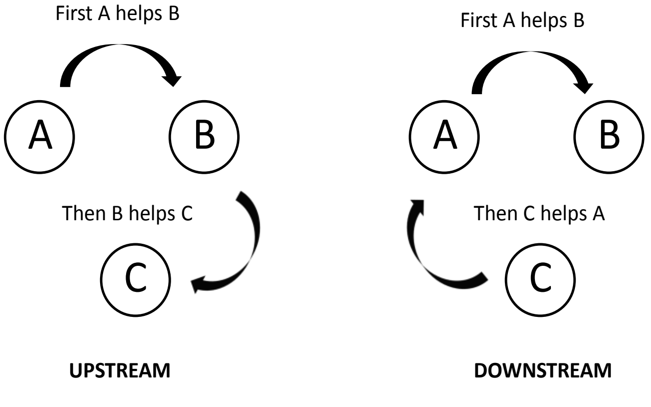
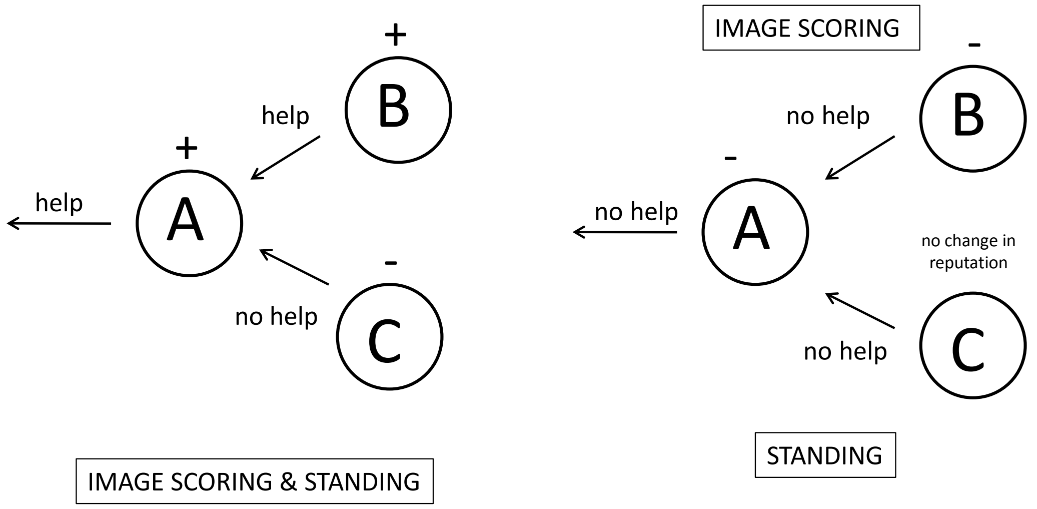


Figure 2.1 Upstream (based on positive experience) and downstream (based on reputation) indirect reciprocity (adapted from Nowak and Sigmund, 2005).

IR can be viewed from the perspective of punishment and reward. Individuals of low cooperative reputation are punished by being denied help or altruistically (at one's own cost) rewarded by being given help (Fehr and Fischbacher, 2003, Rand et al., 2009). It has been proposed that rewarding in IR is, at least partly, truly altruistic in the sense that it occurs even if it does not entail reputational benefits (Engelmann and Fischbacher, 2009). Cooperation in IR can have various motives: an individual can help either to reward someone of a high cooperative reputation or in order to increase his own reputation. Engelmann and Fischbacher (2009) disentangled them by making participants' reputations private or public. When donors' reputation was public the helping rate increased in comparison to when it was private, but even in the private condition cooperation reached a relatively high level (22%-49% - private helping rate, 66%-86% - public helping rate). The researchers interpreted this as evidence for 'pure' or altruistic IR, not contaminated by reputational concerns and motivated by the pure willingness to punish non-cooperators by not helping them and reward cooperators by providing them costly help. However, another result slightly undermines such a conclusion: participants were willing to help people of unknown reputations at a similar rate to those with known reputations in both public and private conditions. In that case, rewarding cooperation or punishing the lack of it cannot explain cooperative behaviour and it appears that a more general rule is used (e.g. upstream reciprocity or donors optimistically assumed that although unknown, the recipient's reputation is high). The concept of pure IR contradicts the original definition of Alexander (1987) according to whom individuals' reputations are "*continually being assessed and reassessed by interactants (…) on the basis of their interactions with others*" (p.85). Alexander pointed out the possible benefits coming from IR which are: (1) profitable direct interactions with others who observed a cooperative act (which corresponds to CA), (2) direct compensation from all or part of the group such as the increase of status or (3) sharing the success of the group within which the cooperator acted which contributes to the cooperator's own descendants. In pure IR donor's reputation can neither be assessed nor rewarded by others.

2.4. Conceptual extensions of indirect reciprocity

Before Alexander published his theory of IR, a similar concept appeared in the work of Sugden (1986). Sugden described a game in which individuals use someone's 'standing' in order to decide whether to cooperate with them or not. Good standing is achieved by cooperating with individuals of good standing and reputation does not decrease when help is refused to those of bad standing. Hence standing is a way of measuring cooperative reputation in which indirect tit-for-tat is rewarded. Another method of assigning reputations is ‘image scoring’ where a positive reputation is given to an individual who cooperated in the past while reputation decreases if help is denied even to individuals of low reputational score (see Figure 2.2).



‘+’ increase in reputation, ‘-’ decrease in reputation

Figure 2.2 Reputational consequences of helping and refusing to help individuals of different reputations under image scoring and standing.

Pollock and Dugatkin (1992) compared a strategy of directly reciprocal tit-for-tat (TFT) with Observer TFT, which corresponds to image scoring where the image is based solely on the last move, and found that when the duration of cooperative encounters was uncertain reputation-based cooperation was more likely to emerge than when the probability of future interactions was high. In another model of image scoring individuals had different thresholds of donating help: they would only help those whose image score was equal or higher than their threshold (Nowak and Sigmund, 1998). Simulations revealed the dominant strategy in which individuals helped everyone with a score of at least 0. However, when mutations were added, in the long run, image scoring was vulnerable to the invasion of unconditional cooperators and these in turn were likely to be invaded by unconditional defectors. It was also shown that in order for image scoring to be evolutionarily stable, the benefit to the recipient multiplied by the probability of knowing the recipient’s reputation needs to exceed the cost of the cooperative act (Nowak and Sigmund, 1998).

The mechanism of image scoring does not appear to be realistic because it does not account for justified defections (not helping a non-cooperator). If, as suggested by Fehr and Fischbacher (2003), IR relies on punishing and rewarding individuals of different reputations, punishing a non-cooperator by not helping him should be perceived as fair. The standing strategy avoids this problem and good standers were shown to invade a population of image scorers (Leimar and Hammerstein, 2001). Moreover, when the image scoring model was re-analyzed with the introduction of errors, defectors soon became the most successful group; standing proved to be much more resistant to the invasion of defectors (Panchanathan and Boyd, 2003). Although standing appears to be a perfect candidate for a mechanism people use to assign reputations, an empirical study undermined its applicability. Milinski and colleagues (2001) experimentally distinguished between image scoring and standing and found that participants' behaviour was more compatible with the former. The researchers speculated that, although standing is superior to image scoring in theoretical models, in real life people might not have the cognitive capacity to process such complex assessments. Ideally, under the standing strategy, upon seeing defection, an individual needs to know whether the recipient was seen defecting, but also, assuming the recipient defected in the past, whether the defection was directed to an individual seen defecting and whether that individual defected because of someone else's defection, etc.

The question of what norms describe whether someone is perceived as good or bad was addressed in three theoretical papers (Ohtsuki and Iwasa, 2004, Ohtsuki and Iwasa, 2006, Ohtsuki and Iwasa, 2007). An examination of all possible ways in which reputation can be assigned revealed that there are eight strategies ('leading eight') which, when in use, can maintain the evolutionary stability of IR (Ohtsuki and Iwasa, 2004). The common features of the leading eight heavily rely on actions towards a good person: helping a good person is always good and not helping a good person is bad. Moreover, not helping a bad person is good. Among the leading eight are standing and 'judging' – a strategy similar to standing in which helping a bad person is considered as bad. All of the leading strategies discriminate between justified and unjustified defection. Despite the theoretical advantage of standing-like strategies, the problem still remains how people assess reputations in reality. The troublesome gap between theoretical models and Milinski et al.'s (2001) study has not yet been explained but a number of studies showed that humans take into account second-order behaviour of others when rewarding or punishing them. Although punishing a non-cooperator at one's own cost benefits the group, people are not willing to impose second-order punishment of non-punishers or second-order reward of punishers (actually, they are likely to punish the punishers, see Herrmann et al., 2008). People do, however, support rewarders of cooperators and punish non-rewarders, which suggests that they willingly react to positive sanctions (Kiyonari and Barclay, 2008). On the other hand, it has been reported that people who engage in altruistic punishment are desired as interaction partners and are entrusted more money than non-punishers (Nelissen, 2008). If cheating or not cooperating is viewed as a form of punishment then humans, in contrast to the assumptions of image scoring, do seem to consider the context of others' non-cooperation. Moreover, people treat punishers who incurred various costs of punishment differently which suggests a complex reputation assessment technique (Nelissen, 2008). Finally, when discussing which of the two models image scoring or standing is more realistic, it is worth considering how strict the assumptions of IR models are. In modelling IR individuals cannot re-meet, whereas no one can prevent individuals from interacting again in a natural environment. This artificial feature of modelling IR has been addressed by Roberts (2008) who showed that when re-meeting is possible standing outperforms image scoring.

In a standard IR model, acts of cooperation increase reputation whereas not giving to another person when there is a chance to do so, decreases it. More specifically, in image scoring giving results in +1 score and non-giving in -1. Image scoring, then, assumes opposite but equal values of good and bad deeds. It has been theoretically shown that the value of an act depends on the benefits of cooperation (Rankin and Eggimann, 2009). High benefits of cooperation favour the evolution of a judgement bias in which cooperation has more value than non-cooperation whereas with small benefits of cooperation judgement bias is shifted towards non-cooperation. Rankin and Eggimann (2009) point to the studies that suggest judgement bias in humans by reporting that cheaters are remembered better than cooperators (e.g. Vanneste et al., 2007). Judgement bias has also been noticed in the area of book reputations: negative reviews affected the selling rates much more than positive reviews (Chevalier and Mayzlin, 2006). Given the weak theoretical underpinnings of image scoring, in this thesis, I will be referring to IR as a function the standing strategy.

2.5. Empirical evidence for indirect reciprocity

In recent years an abundance of studies supported the basic premise of IR and CA, namely that caring about reputations affects human behaviour. Reputations can be formed whenever an individual is observed by others, hence, according to reputation-based cooperation theories, in such a context individuals should be more cooperative than when alone. Indeed, even under conditions of anonymity, cues of being watched make people more cooperative. Presenting participants with eyespots on the computer screen increased their generosity (Haley and Fessler, 2005) and displaying an image of eyes next to an honesty box instead of a neutral image resulted in higher contributions (Bateson et al., 2006). Similarly, people exhibited higher levels of cooperation when they were "watched" by an image of a robot (Burnham and Hare, 2007). However, in a carefully controlled experiment, cooperative behaviour of participants working in complete anonymity in a private context did not differ from when participants were anonymous but worked in the same room with others (Lamba and Mace, 2010). This finding suggests that people can recognize when they are really anonymous even if they are surrounded by others (public context) and act strategically i.e. display lower levels of cooperation than when their behaviour is revealed. The differences between the aforementioned studies might have been due to the economic game used. In the Ultimatum game employed by Lamba and Mace (2010) sharing resources with another person does not reflect true altruism because the recipient can punish the proposer if the split is unfair. Hence, under both private and public conditions it is in the proposer's interest to offer a split fair enough to be accepted by the recipient. Supportive of such an explanation are the results obtained by Charness and Gneezy (2008) who showed that revealing someone's family name increased their generosity in the Dictator game (where the recipient is completely passive with respect to how the money is shared) but not in the Ultimatum game.

Human cooperative behaviour is affected even when reputation gained lasts for a very short time. When reputations can be transferred from one economic game to another one, contributions to the public pool increase (Semmann et al., 2004). For enhancing cooperation it is important that identification information (e.g. a photograph) is made public together with the cooperative decision: neither displaying photographs alone nor cooperative decisions changes participants' behaviour (Andreoni and Petrie, 2004). When given a choice people prefer to reveal their donations rather than keep them anonymous and the possibility of choice between the two options increases the overall generosity (Andreoni and Petrie, 2004). By avoiding confidentiality individuals ensure that they are not perceived as selfish which might occur if they donated anonymously. Also, in a real-life setting participants were more willing to provide assistance to strangers when their pro-social offers were made in the presence of their peers than when they were anonymous (Bereczkei et al., 2007).

A broad assumption of IR is that people are concerned about their cooperative reputations when observed by others, but in particular when others have a chance to indirectly reciprocate their cooperation. Engelmann and Fischbacher (2009) reported that in their sample only 20% of participants were categorized as non-strategic i.e. helped other individuals even though it did not affect their reputation whereas the behaviour of over 50% was classified as strongly strategic i.e. they increased their helping rate at least twice when that could be reflected in their reputational score in comparison to when their reputational score was not made public. This result ties in with the notion of different cooperative types in humans. In a sample investigated by Kurzban and Houser (2005), reciprocators (i.e. those who reacted to others' behaviour) made up 63% of players whereas unconditional cooperators and free-riders constituted a much smaller proportion of participants. Strategic vs. non-strategic IR was investigated in a study on heterogeneous social preferences (Simpson and Willer, 2008). Using an ecologically valid measure of pro-social preferences, the Social Value Orientation Scale, the researchers divided participants *a priori* into egoists and altruists. It was demonstrated that only egoists were affected by the reputational incentives in IR; altruists were willing to split their allocation in a similar way irrespective of whether their decision would be revealed to another splitter or not. Moreover, egoists were also more sensitive to the context in which their partners split the money previously and gave more to those whose reputation was built in a private rather than public condition. Hence, they regarded helping done without realising that others would observe it as more reward-deserving than helping in the context of reputational incentives.

Another intrinsic feature of IR is that cooperators reap benefits from cooperation in the form of indirectly reciprocated acts of kindness. Indeed, in an experimental setting, participants of high image score received money more frequently than those with lower image score from others who observed their generosity (Wedekind and Milinski, 2000). In another experiment, it was noted that 48% of donors considered the recipient's reputation when making a decision about their donation and tended to help those with high reputational score (Seinen and Schram, 2006). At the same time 35% of donors made their cooperative decision taking into account their own cooperative score (the lower the score the more likely they were to provide help). The high percent of donors basing their cooperative decisions on the reputational status of the recipient resulted in the individuals' concern over their own reputation.

Giving to a charity has also been shown to yield cooperative rewards (Milinski et al., 2002a). Participants who decided to donate more money to UNICEF received more from others in the subsequent rounds. Moreover, the sum of the donations to UNICEF and other players was positively associated with votes in the election for the students' council. In another study, high cooperative reputation translated to higher payoffs in directly reciprocal interactions (Wedekind and Braithwaite, 2002). Further evidence for benefits that can be gained from reputation comes from a controlled field experiment on EBay. An EBayer with an established reputation received a significantly higher price for his items than a new seller (Resnick et al., 2006). In this example however, IR is probably not the only mechanism affecting people's behaviour. EBayers entrust their money to sellers expecting in exchange high-quality goods. Hence such an interaction can be viewed as directly reciprocal.

Theoretical models showed that systems of IR can be stable under certain conditions (e.g. Nowak and Sigmund, 2005). The question remained, however, whether in reality cooperation would be maintained when individuals have a chance to cooperate with others of known reputations. Cooperation in social dilemmas i.e. situations when one can either perform an action that benefits oneself or an action that benefits others has been extensively investigated in experimental economics. The unanimous finding is that, under conditions of full anonymity, although individuals show some level of pro-sociality at the beginning of the game, with repeated rounds, cooperation drops dramatically and the game finishes with a majority of selfish actors (e.g. Isaac et al., 1994). Such a result can be viewed as a "tragedy of the commons": if individuals have free access to a public resource, they start to selfishly overexploit it and the resource is quickly depleted (Hardin, 1968). Punishment was demonstrated to be an effective way in which cooperation in social dilemmas is sustained (Fehr and Gächter, 2002); however, recent reports on antisocial punishment question its role in enhancing cooperation (Herrmann et al., 2008). Milinski and colleagues (2002b) addressed a question of whether the IR context could also prevent individuals from becoming selfish. In their experiment, social dilemma rounds were alternated with rounds in which individuals could transfer money to a third-party without the possibility of direct reciprocation. The level of cooperation remained stable over rounds. In another condition, participants first played a number of social dilemma rounds, after which IR rounds were introduced and played subsequently. After the initial decrease in cooperation during rounds without reputational incentives, cooperation recovered and reached a high level. Moreover, when the payoffs of groups with IR-alternated and non-alternated social dilemma rounds were compared, groups with IR earned significantly more money. In summary, it was shown that IR, like altruistic punishment, works well at maintaining high levels of cooperation. Considering that many real-life situations resemble social dilemmas and that people make use of reputations, such a design accurately reflected a possible mechanism behind the stability of human cooperation (Milinski et al., 2002b).

2.6. Theories behind reputation-based partner choice

As presented above, IR potentially accounts for instances of cooperation which could not be explained by traditional theories such as kin selection and reciprocal altruism. However, a problem remains how to treat unconditional cooperation such as donations to a third party not based on their reputational score, which frequently occur in human societies (e.g. giving to charity or donating blood). Inspiration derived from theories of costly signalling and biological markets resulted in a concept of another mechanism of human cooperation which applies to unconditional helping – competitive altruism (Roberts, 1998).

The core assumption of CA is that a cooperative act functions not only as a mere action in response to another individual's cooperation but as a signal informing about some underlying qualities of the signaller. Such an interpretation stems from comparing cooperation to costly displays in the handicap principle (Zahavi and Zahavi, 1997). An individual's behaviour can serve as an honest signal of quality to others if it fulfils certain criteria (Smith and Bliege Bird, 2005). First, the signal needs to be (at least temporarily) costly to the sender, a condition which is by definition satisfied by a cooperative act. In order for the signal to be reliable, the relative cost of it has to vary between individuals of different classes i.e. it should be less affordable to individuals of low status. The signal needs also to be broadcasted to the audience so it should not be performed in anonymous conditions. Finally, both the sender and the receiver of the signal need to reap benefits. In a typical handicap, e.g. a peacock's tail, the receiver of the signal benefits only by acquiring ecologically important information about the sender. With cooperation, if the receiver of a signal is also the receiver of a cooperative act, the benefit increases considerably. Signal senders, on the other hand, benefit by acquiring access to valuable social and mating partners. As cooperation meets all the above requirements (see Table 2.1) it is tempting to view it in the context of costly signalling. Zahavi (1995) did notice the potential in the handicap principle to extend to cooperation; however, he did not develop the theory fully. It is worth noting that Zahavian handicap principle is just one way through which a signal can be reliable e.g. according to Brown and Moore's (2002) classification, expression of pro-social emotions would be a physiologically constrained honest index of cooperation because it involves facial movements which are not under the voluntary control.

Costly signalling via cooperation can occur if individuals vary in quality and when quality can be reflected in behaviour. Gintis et al. (2001) developed a multi-player game in which they examined whether costly signalling can increase in a population when it is initially rare. The researchers found that costly signalling of quality in the form of contributions to the public pool can be evolutionarily stable if the signalled qualities are beneficial to potential social and mating partners. Lotem et al. (2002) investigated how individual variation in quality and the introduction of signalling benefits affect cooperation. The main assumption of their model is a positive association between altruism and quality, so high-quality individuals provide extensive help whereas low-quality individuals tend to defect. It was shown that, when quality is signalled, unconditional altruism by high-quality individuals can emerge and stabilize the system in which low-quality individuals either defect or reciprocate (Lotem et al., 2002).

Table 2.1 Examples of human and animal cooperation functioning as a handicap.

Requirement	Animal example (allopreening)	Human example (charity giving)
costly	expenditure of time and energy	expenditure of resources
reliable	more expensive to low-quality individuals who sometimes cannot afford it	more expensive to low-resource individuals who sometimes cannot afford it
observable	publicly available to observe	either publicly announced by the charity or advertised in the form of a charity badge
benefits the sender	reputation of a high-quality individual resulting in better access to mates and coalitions	reputation of a high-quality individual resulting in better access to mates and coalitions
benefits the receiver	removal of parasites and maintenance of plumage condition, information about the quality of potential coalition partners	improving the welfare of the group in need, information about the resources and pro-sociality of the signaller

Anthropological literature gives accounts of pro-social behaviours likely to function as costly signals. Skilled men from the Meriam tribe in Melanesia engage in dangerous turtle hunting for reasons clearly different from acquiring food for themselves (Bliege Bird et al., 2001). Hunters widely distribute the meat among group members during a feast. The majority of attendees know the identity of the successful hunters and such recognition results in social and reproductive benefits (Smith et al., 2003a). Another example comes from indigenous peoples of the Pacific Northwest Coast who organize potlatches, feasts during which they make gifts to members of nearby villages in order to outperform them in generosity (Jonaitis, 1991). It is still under debate whether pro-sociality really translates to some underlying qualities or 'good genes'. Experimental research linked altruism to intelligence, a highly desirable quality in both social and mating partners (Millet and Dewitte, 2007). The IQ of participants classified as altruists was significantly higher than those classified as mere cooperators (giving a fair share) or egoists. Brown et al. (2005)

found a link between altruism and health in older adults; however, the causality of this relationship could not be determined. More research is needed to explore the question whether pro-social behaviour signals characteristics other than wealth and cooperative intentions.

Apart from costly signalling, another cornerstone of CA is competition for partners, which also plays a crucial role in the biological market theory (BM). Although when first describing CA Roberts (1998) did not identify the link between the two concepts, other researchers noticed and pointed it out later (Chiang, 2010, Barclay, 2010). According to BM forces of supply and demand observable in economic markets apply equally well to animal and human ecology (Noë and Hammerstein, 1994, Noë and Hammerstein, 1995). Whenever an asymmetry in the possession of certain commodities exists in a population, individuals benefit from the exchange of different types of commodities. The rules of supply and demand determine the value of the commodities and the value of individuals possessing different commodities as social partners. Partner choice and competition for the most desirable partners constitute an essential ingredient of BM. A common feature of costly signalling and BM is that individuals can advertise their commodities or qualities as exchange partners, but unlike costly signalling, BM acknowledges the role of dishonest advertisements. Another feature of BM is that the market is based entirely on reputation, excluding theft or forceful acquisition of commodities.

Both CA and BM assume that in cooperation markets individuals should show off pro-social tendencies in order to acquire access to the most cooperative partners. In a recent paper Chiang (2010) contrasted the characteristics of BM and CA (see Table 2.2) and presented evidence for the quality effect of partner selection. Self-interest motivated participants to preferentially select generous sharers if they were the recipient of a share and tolerant recipients if they were the sharer. Moreover, it was theoretically shown that free market of partner choice is just one requirement to be met for the emergence of fairness. Another crucial aspect affecting cooperation is what partner preferences individuals actually have (Chiang, 2010). Reputation-based partner choice can also be viewed from the perspective of Fisherian runaway selection. In the original model, evolution of some sexually attractive traits is based on positive feedback (Fisher, 1930). A preference for a specific trait in the opposite sex makes this trait advantageous and stimulates its proliferation. Human pro-social behaviour might have been shaped by social runaway selection if the preference to form partnerships with the most cooperative individuals was strong enough to induce evolutionary pressure (Nesse, 2007).

Table 2.2 Differences between some of the assumptions of CA and BM (adapted from Chiang, 2010)

Difference	Competitive Altruism (CA)	Biological Market (BM)
Assessment stage	Every individual is assessed by others	Sampling only a set of potential partners, sampling techniques affect assessment
Cheating	Does not pay off in the long run, so should not undermine reputation-based systems	Can impede market mechanism
Partner selection	Quality effect: the most altruistic individuals preferentially interact with each other	Quantity effect: when choosers outnumber potential partners, the partners have an advantage in the form of interacting with many choosers and *vice versa*

2.7. Competitive altruism: theoretical and empirical evidence

In the original paper Roberts (1998) proposed a way in which CA or reputation-based partner choice could function. In the first stage, 'assessment', individuals make cooperative displays and evaluate each other's quality as a partner. In the second, 'partnered' stage individuals choose a partner and if accepted interact with them. There are several assumptions behind CA: (1) individuals need to vary in quality as partners; (2) individuals' behaviour has to be public and (3) individuals should be able to choose partners. If cooperation reflects partner quality, the most cooperative individuals are also most desired as partners, and hence, can be more selective when deciding who to interact with. Competition for the most cooperative partners results in the escalation of cooperation with individuals trying to reputationally outperform each other. Little theoretical work has been done on CA. Evolutionary simulations showed that the stability of cooperation under conditions similar to CA (where the dismissal of uncooperative partners is possible) depends on the composition of behavioural types in a population which results from mutations (McNamara et al., 2008). In this model selfishness was discouraged by the threat of being rejected as a partner. Barclay (2010) found that a population of competitive altruists can invade a population of non-cooperators provided that some individuals are at least initially capable of providing non-costly help.

As Roberts (1998) stated *'altruism could persist (…) through competition: competition for the attentions of other altruists, competition for mates'* (p.429). Cooperation as a signal can be attractive to potential social and sexual partners. In the latter case, generosity is advertised in order to impress attractive mates. It was demonstrated that more cooperative individuals are perceived as more attractive and that more cooperation is directed towards more attractive than less attractive individuals (Farrelly et al., 2007). Men were also shown to contribute more money to charity when observed by a member of the opposite sex than of the same sex (Iredale et al., 2008). With cooperation serving as a mating signal it is difficult to disentangle the different motives behind choosing a cooperative partner and even more difficult to estimate any long-term benefits associated with cooperative displays. The framework with individuals cooperating in order to impress social instead of mating partners proved to be more fruitful with regard to producing evidence for CA.

Barclay (2004) designed three experimental conditions in which participants' decisions in a social dilemma game preceded (1) an unknown game with other players, (2) a game in which it was desired that other players considered the target individual as trustworthy (3) the same game based on trust in which additionally participants could choose a partner they wanted to play with. Cooperative contributions were higher in the two reputational conditions (in which players knew that the other game would be based on trust) than in the no-reputation condition. Barclay's study also proved that incentives behind reputation building can stabilise cooperation. Contributions to the public pool did not drop, as usually happens with repeated games, but remained high over a few rounds thanks to participants' willingness to build up reputation. Hardy and Van Vugt (2006) examined the association between individual cooperation and the status in a group. Political esteem and social prestige of individuals were positively related to their pro-social contributions. Cooperators were also preferred as exchange partners over non-cooperators. In another study, by varying the incentive to cooperate (anonymous contribution, public contribution and public contribution with partner choice) it was demonstrated that participants' contributions were related to the motivation to gain cooperative reputation (Barclay and Willer, 2007). Apart from showing that the most cooperative participants were chosen most frequently as partners, the results suggest that people are sceptical of cooperative signals produced in a condition with potential high reputational benefits. Although Barclay and Willer's (2007) study demonstrated the benefits of reputational generosity in the form of easier access to desired partners, the researchers failed to find any other advantages of it. Using different methodology Sylwester and Roberts (2010) observed that not only do the most cooperative players acquire the socially desirable individuals as partners but that they also receive higher returns from those partners. Thus, the short-term cost of investing in reputation is recouped in the long-term by engaging in profitable interactions.

2.8. Conclusions

In recent years, two lines of research have been developing independently despite tackling the same topic of reputation-based cooperation in humans. The two investigated theories, IR and CA, were both conceived by biologists but the majority of empirical research on IR has been conducted by behavioural economists whereas CA has been mostly the domain of evolutionary and social psychologists. IR has received disproportionately more attention than CA even considering the fact that it was formulated earlier than CA. To provide a crude measure: a search in the Scopus database yields 120 entries for IR and only 13 for CA. This is puzzling given that CA appears to be a more parsimonious and robust mechanism for explaining cooperation than IR. As argued by Bshary and Grutter (2006), when viewed from the perspective of communication network theory IR is a complex mechanism requiring at least two conditions. First, observers of cooperative acts need to gain some personal benefits from the information collected (e.g. finding cooperative partners), and second, cooperators need to gain from access to the observers. Interestingly, the two necessary assumptions constitute the core of CA. In both mechanisms, IR and CA, individuals cooperate in order to be seen as cooperators by others, but in IR their cooperative behaviour results in a one-off benefit from the third party whereas in CA it results in access to socially desirable partners.

There are a number of possible definitions of IR which I presented in section 2.3. According to Alexander (1987), in its broadest sense, IR may even involve the same mechanism of partner choice as CA: *"Indirect reciprocity must have arisen out of the search for interactants and situations by which to maximize returns from asymmetrical, hence highly profitable direct social reciprocity."* (p.97). Any form of non-kin cooperation depends upon a correlation between giving and receiving - cooperators must be more likely to be recipients than non-cooperators. For example, direct reciprocity can only work if it is contingent on the cooperative response of a recipient. Otherwise non-cooperators would keep increasing their profits by accepting cooperation without reciprocating and in the evolutionary perspective would oust cooperators. The easiest way to define IR is by an analogy to direct reciprocity. Again, it is only possible for it to work if there is a discrimination - in this case, cooperators must cooperate preferentially with those who have cooperated with others (image scoring, where individuals cooperate indiscriminately to boost their reputation, is unstable, Leimar and Hammerstein, 2001; Panchanathan and Boyd, 2003). CA is more robust than IR because it accounts for unconditional cooperation. Where CA is different is that it invokes two stages in each of which a different social interaction occurs. While in IR individuals must cooperate discriminatingly, in CA, cooperation may be indiscriminate in stage 1 because there is no requirement within the stage 1 framework that cooperators receive more than non-cooperators. Instead cooperators receive their payback in the separate stage 2 based on dyadic interactions with chosen partners.

IR and CA both aim to explain the extensive cooperation towards unrelated individuals observed in humans using a reputation-based framework. Building and assessing reputations (in particular second- and higher-order) is a cognitively complex activity unlikely to be found in non-human animals. While IR received much more publicity than CA, it does not address the phenomenon of unconditional cooperation i.e. cooperation towards individuals of no or low reputations. Assessment of individuals in IR constitutes an unsolved problem which does not exist in CA. Moreover, the strict assumptions of IR seem artificial when applied to human societies where reputation transmission often facilitates long-term partnerships and individuals can interact repeatedly with the same partner. Although, unlike IR, CA has not gained much attention in terms of theoretical modelling, it has a strong conceptual basis in the form of biological markets and costly signalling theories. Despite many similarities in IR and CA, researchers publishing papers on one of the two theories do not typically address the other one (with the exception of Barclay, 2004, Sylwester and Roberts, 2010). There is a need for an empirical and theoretical evaluation to examine which of the two theories makes a more likely mechanism for interpreting human cooperation. Ideally, the two mechanisms should be contrasted in one model or study, so that the effects of the two contexts on behaviour could be distinguished and measured.

3. Profitable cooperation through competitive altruism

Explaining unconditional cooperation, such as donations to charities or contributions to public goods, continues to present a problem. One possibility is that cooperation can pay through developing a reputation which makes one more likely to be chosen for a profitable cooperative partnership, a process termed competitive altruism (CA) or reputation-based partner choice. We tested whether people exhibit higher levels of cooperation in the presence of reputational concerns and whether investing in reputation pays off. Participants played a public goods game (PGG) with varying reputational incentives followed by another cooperative game for which they could sometimes choose partners. Participants contributed significantly more in the first stage of the game, when they knew that their contributions would be revealed to other players in stage two. Moreover, the contributions were even higher when participants were told that they would be able to choose partners for the second game. Reputational competition was strongest when it was possible for participants to receive a higher payoff from partner choice. Most importantly, we showed for the first time that investing in a cooperative reputation can bring net benefits through access to more cooperative partners. We concluded that CA provides an alternative to indirect reciprocity (IR) as an explanation for reputation-building behaviour. Furthermore, while IR depends upon individuals giving preferentially to those of good standing, CA can explain unconditional cooperation.

3.1. Introduction

Research on cooperation has shown that people are more generous when they are watched by someone or even when they are exposed to images of eyes (Bateson et al., 2006, Bereczkei et al., 2007). The rate of cooperation increases also when the identity of the individual is revealed (Andreoni and Petrie, 2004). Considering these findings, generosity appears to be a context-dependent behaviour expressed in the presence of reputational incentives.

One reason why it might pay to be seen to cooperate is indirect reciprocity (IR; Alexander, 1987). Experiments have shown that people do indeed prefer to help those who help others (e.g. Milinski et al., 2002b). Moreover, investing in reputation pays in the long run: despite the initial expense, individuals benefit by receiving more cooperation from others in subsequent rounds of IR (Wedekind and Braithwaite, 2002). However, problems remain with IR as a general explanation for reputation building (Roberts, 1998). In particular, IR depends upon cooperation being directed to cooperative individuals, so cannot explain displays of unconditional cooperation.

An alternative theory for reputation formation is that of competitive altruism (CA; Roberts, 1998). This theory stresses the role of partner choice for profitable relationships and is based on a two-stage process in which individuals first have a chance to build up reputations through making generous displays and secondly choose partners for further interactions. CA postulates that individuals seek to acquire the best cooperators as partners. According to biological market theory such cooperative pairing will be assortative (Noë and Hammerstein, 1994). In CA, the benefit of a high cooperative reputation, gained through being accepted as a partner, is the return from another cooperative individual in a dyadic interaction.

In support of CA, research has shown that people were more cooperative when they expected to play a dyadic trust game with a chosen partner later than when they knew they would not be able to choose a partner or when they did not expect to play a further game (Barclay, 2004). Another study demonstrated that participants contributed more when their contributions were to be revealed to others than when they remained anonymous. It was also found that status and social prestige increased in proportion to donations made to the group (Hardy and van Vugt, 2006). Finally, by varying whether contributions were anonymous or public and whether participants had a choice of partner, Barclay and Willer (2007) demonstrated that participants' contributions were positively related to the motivation to gain cooperative reputation. The study also provided evidence for a preference for the most cooperative players. However, none of these studies has demonstrated net monetary benefits from investing in a cooperative reputation.

Here we investigate the benefits coming from reputation in the form of partner choice and payoffs from interactions with partners. The study also involves varying the potential gains to be made from a partner to test whether we find an increase in contributions when reputation building is followed by higher potential rewards.

We conducted two experiments referred to as Study 1 and Study 2. Both studies consisted of two stages in which participants first had an opportunity to build up reputation (Stage 1) and could then to a lesser or greater extent (depending on the condition) make use of the information about other players' reputations e.g. choose partners for further interactions (Stage 2). Study 1's aim was to demonstrate that (1) public information and (2) partner choice increased cooperative reputation-building behaviour. Our findings were consistent with those of earlier studies including Barclay & Willer (2007) confirming that we used the appropriate methodology. Study 2 had a different payment structure and participants could always choose partners for games with two reward levels.

3.2. Study 1

3.2.1. Method

Participants

15 groups of four participants (48 women, mean age = 21.04, SD = 2.33 and 12 men, mean age = 24.92, SD = 4.89) were recruited from Newcastle University. The groups consisted either of females (seven groups) or were mixed (six groups with three females and one male and two groups with three males and one female). Participants were rewarded with money in such a way that the highest earner in a group of four received £20 and others received nothing. An informed consent was obtained from all participants and they were debriefed after the experiment.

Design

Participants played public goods games (PGGs) in a 2-stage within-subjects design. In each round participants were endowed with 10 lab pounds. Participants could either keep this money, or contribute all or some part of this money to a common pool. The sum of the contributions was doubled and shared equally among the players irrespective of how much they contributed (each invested pound yielded only 50p to the investor). In Stage 1 participants played PGGs in a group of four whereas in Stage 2 they played it in pairs with a person they chose or with an arbitrarily assigned person. Participants' contributions from Stage 1 were not revealed to other players until Stage 2 took place. Because of this, participants' decisions in Stage 1 were only influenced by the four experimental conditions. In the *Anonymous* condition round of Stage 1 participants were told that the contribution that they made would not be revealed to others at all. In the *Public* condition round participants learned that what they contributed would be visible to other players before one round of Stage 2 of the game. The *Public* condition represents IR framework in which cooperation can be indirectly reciprocated by cooperating with an individual of h igh cooperative reputation. In the *Choice* condition round participants were informed that what they contributed would be revealed to other players in Stage 2 before playing a round for which they would be able to choose partners. Finally, the *Choice-Bonus* condition round differed from the *Choice* condition by the fact that in the former participants' contributions were to be revealed before playing a special bonus round in which the amount put in the pool was multiplied by eight and evenly distributed to participants.

Table 3.1 Four experimental conditions and the incentives to give money in each.

Condition	Stage 1 contributions revealed to others	Possibility of choosing a partner in Stage 2	High-stake game in Stage 2
Anonymous	no	no	no
Public	yes	no	no
Choice	yes	yes	no
Choice-Bonus	yes	yes	yes

Hence, in Stage 2 there were two reward levels: in the *Anonymous, Public* and *Choice* conditions the amount in the pool was multiplied by two before distributing it to participants while in the *Choice-Bonus* condition it was multiplied by eight. In both cases, although the game retained the form of the PGG, there was no social dilemma. In the three conditions with standard payoff each invested pound yielded exactly £1back so an individual's gain was only affected by their partners' contributions. In the *Choice-Bonus* condition the rational choice was to contribute the whole allocation to the pool because each individual benefited from their own contribution (each invested pound yielded £4). In the *Choice-Bonus* round the incentive to acquire a cooperative partner was greater because if both partners invested everything the final profit could reach four times that of the *Choice* condition (see Table 3.1). Although, in the *Choice-Bonus* round, investing everything in the public pool was the most payoff-maximizing decision, we expected a variation in contributions. Previous studies have shown that people tend to behave imperfectly or competitively even if it is in their own interest to cooperate (Kümmerli et al., 2010, Kurzban and Houser, 2005). Hence, the unusual structure of the PGG in the *Choice-Bonus* round still provided room for reputation-building behaviour.

Procedure

Upon arrival, four participants were led to computers separated by screens. Anonymity was also ensured by providing participants with experimental nicknames, so that any reputation was assigned to nicknames and not to the real names of participants. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007). At the beginning, participants read the instructions, in which the nature of Stage 1 and Stage 2 was explained to them, and were familiarized with the game by playing one trial round in which they all were asked to contribute £5. After that, participants played Stage 1 which consisted of four rounds of PGG, each round representing a different condition (see Table 3.2 and Appendix A). The order of conditions in Stage 1 was balanced across participants, in such a way that in each round each participant played a different condition. In Stage 1 participants did not receive any feedback on how much money other players contributed.

Table 3.2 An example of the order in which PGGs were presented. The order of presenting incentives in Stage 1 was balanced across participants. In Stage 2 the conditions were presented in the same order for all four participants (P1-P4), but this order was balanced between the groups (not shown in the table). Cell entries in Stage 1 refer to the information participants received before contributing to PGG whereas cell entries in Stage 2 refer to the way in which contributions from Stage 1 were used.

	Stage 1 (group games)				**Stage 2 (paired games)**			
	Round 1	Round 2	Round 3	Round 4	Round 1	Round 2	Round 3	Round 4
P1	Contribution not revealed in *Anonymous*	Contribution revealed in *Public*	Contribution revealed in *Choice*	Contribution revealed in *Choice-Bonus*	*Anonymous*	*Public*	*Choice*	*Choice-Bonus*
P2	Contribution revealed in *Public*	Contribution not revealed in *Anonymous*	Contribution revealed in *Choice-Bonus*	Contribution revealed in *Choice*	*Anonymous*	*Public*	*Choice*	*Choice-Bonus*
P3	Contribution revealed in - *Choice*	Contribution revealed in *Choice-Bonus*	Contribution not revealed in *Anonymous*	Contribution revealed in *Public*	*Anonymous*	*Public*	*Choice*	*Choice-Bonus*
P4	Contribution revealed in *Choice-Bonus*	Contribution revealed in *Choice*	Contribution revealed in *Public*	Contribution not revealed in *Anonymous*	*Anonymous*	*Public*	*-Choice*	*Choice-Bonus*

After Stage 1, participants were asked to complete a short questionnaire which allowed the experimenter to transfer their contributions to paper. In Stage 2, participants played four rounds of PGG and before each round they received a paper slip from the experimenter which either informed them that no contributions are revealed and no partner choice is possible, contained players' contributions from one round of Stage 1 but partner choice was not allowed, contained contributions and asked them to choose the player they would like to be paired with in a game with a standard gain or a bonus gain. Participants knew that they would be allowed to play with the chosen person only if the person chose them as well; otherwise, their partner would be arbitrarily assigned. However, participants did not learn whether they obtained their desired partner or not. The conditions in Stage 2 were presented in the same order for all participants in a group but their order of presentation was balanced across groups. The results of Stage 2 are not analysed as they do not provide answers to our research questions. P values are two-tailed throughout the thesis and all the parametric and non-parametric tests were used according to whether the data met the assumptions of normality. In some cases (where stated) transformations have been use in order to enable parametric tests.

3.2.2. Results

Friedman ANOVA showed that contribution levels in Stage 1 changed significantly over the four conditions, $\chi(3) = 46.23$, $p < 0.001$ (see Figure 3.1).

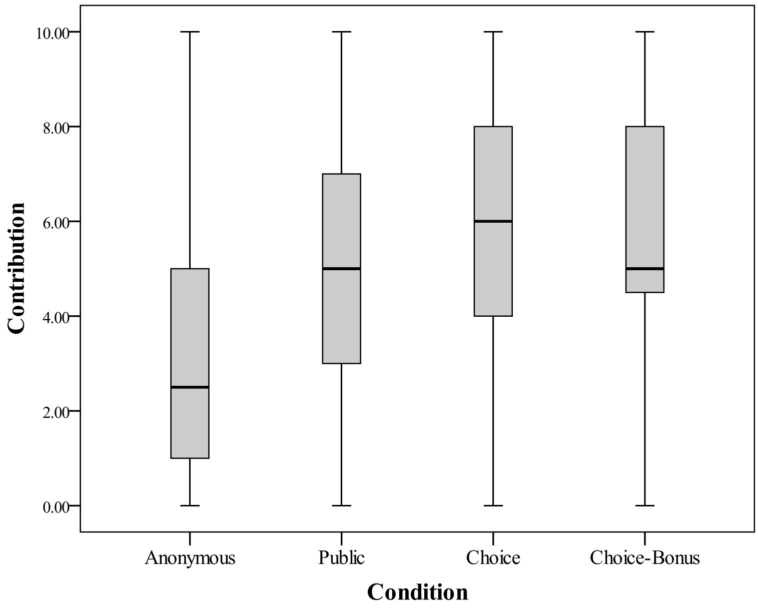


Figure 3.1 Boxplots presenting median contributions (with quartiles and extreme values) to the common pool in lab pounds by condition.

Wilcoxon signed ranks tests revealed that participants contributed significantly more in the *Public* condition (Med = 5, IQR = 4) than in the *Anonymous* condition (Med = 2.5, IQR = 4), z = -3.48, $p < 0.04$. Contributions in the *Choice* condition (Med = 6, IQR =4) were significantly higher than in the *Public* condition, z = -2.83, $p < 0.04$, but did not differ from contributions in the *Choice-Bonus* condition (Med = 5, IQR = 3.75), z = -0.55, p = 0.58. Contributions in the *Public* condition were lower than contributions in the *Bonus* condition z = -2.93, $p < 0.04$. Levels of significance were adjusted using Keppel's modified Bonferroni corrections (see Keppel & Wickens, 2004).

3.3. Study 2

3.3.1. Method

By rewarding the highest-earner with £20 in Study 1 we intended to encourage people to take part in this study and to create a competitive environment. Such a payment method could, however, interfere with the incentive to cooperate (Andreoni, 1995) since participants were rewarded for earning more than others in their group thereby creating in-group competition (see West et al., 2006). We therefore recruited ten groups of four participants, 14 men (mean age = 24.93, SD = 4.32) and 26 women (mean age = 23.38, SD = 3.33) for the second study in which participants received money in proportion to what they actually earned during the game.

Table 3.3 An example of the order in which PGGs were presented. Cell entries in Stage 1 refer to the information participants received before contributing to PGG whereas cell entries in Stage 2 refer to the way in which contributions from Stage 1 were used.

	Stage 1 (group games)				**Stage 2 (paired games)**			
	Round 1	Round 2	Round 3	Round 4	Round 1	Round 2	Round 3	Round 4
P1	Contribution revealed in *Choice*	Contribution revealed in *Choice*	Contribution revealed in *Choice Bonus*	Contribution revealed in *Choice-Bonus*	*Choice*	*Choice*	*Choice Bonus*	*Choice-Bonus*
P2	Contribution revealed in *Choice Bonus*	Contribution not revealed in *Choice Bonus*	Contribution revealed in *Choice*	Contribution revealed in *Choice*	*Choice*	*Choice*	*Choice-Bonus*	*Choice-Bonus*
P3	Contribution revealed in - *Choice*	Contribution revealed in *Choice-Bonus*	Contribution not revealed in *Choice-Bonus*	Contribution revealed in *Choice*	*Choice*	*Choice*	*Choice-Bonus*	*Choice-Bonus*
P4	Contribution revealed in *Choice-Bonus*	Contribution revealed in *Choice*	Contribution revealed in *Choice*	Contribution not revealed in *Choice-Bonus*	*Choice*	*Choice*	*Choice-Bonus*	*Choice-Bonus*

There were two female-only groups, four groups with an equal number of males and females, three groups with three females and one male and one group with three males and one female. In order to investigate in greater depth the most novel conditions which reflected the CA framework, we restricted Study 2 to the *Choice* and the *Choice-Bonus* conditions (see Table 3.3). Study 2 had a similar structure to Study 1 with the only difference that participants played two rounds of the *Choice* condition and two rounds of the *Choice-Bonus* condition. Hence, the total number of rounds was the same as in Study 1. As in Study 1, participants knew that they would only play with their desired partner if this person chooses them as well. We also administered a questionnaire at the end of the experiment asking how participants perceived the strategy they adopted while playing the games.

3.3.2. Results

Participants were ranked within a group according to their contribution (1= top contributor). Participants' rank was significantly negatively correlated with the number of times they were chosen as a desired partner in all four rounds; *Choice* rounds: $r_s = -0.34$, $p < 0.05$ and $r_s = -0.73$, $p < 0.01$, *Choice-Bonus* rounds: $r_s = -0.57$, $p < 0.01$ and $r_s = -0.61$, $p < 0.01$. The ranks of participants who were paired with a chosen partner were significantly higher than of participants who were assigned a partner in all four rounds (Figure 3.2 and Table 3.4).

Further analyses were conducted on the averages of contributions over the two rounds of each condition. Investments made by participants in Stage 1 were strongly and significantly correlated with investments by their partners in Stage 2 ($r_s = 0.59$, $p < 0.01$ for the *Choice* rounds; and, $r_s = 0.57$, $p < 0.01$ for the *Choice-Bonus* rounds; Figure 3.3). Wilcoxon signed ranks tests revealed a non-significant trend suggesting that participants contributed more in the *Choice-Bonus* rounds (Mdn = 7.00, IQR = 3.75) than in the *Choice* rounds in Stage 1(Mdn = 6.00, IQR = 4.38), z = -1.66, p = 0.097.

Table 3.4 Mann-Whitney U tests showing differences in the ranks of participants who acquired their desired partners and those who did not. *p < 0.05, **p < 0.01

Round	U	z	N assigned	N desired
Choice 1	106.500*	-2.396	24	16
Choice 2	43.000**	-3.994	14	26
Choice-Bonus 1	45.000**	-4.213	18	22
Choice-Bonus 2	99.000**	-2.605	24	16

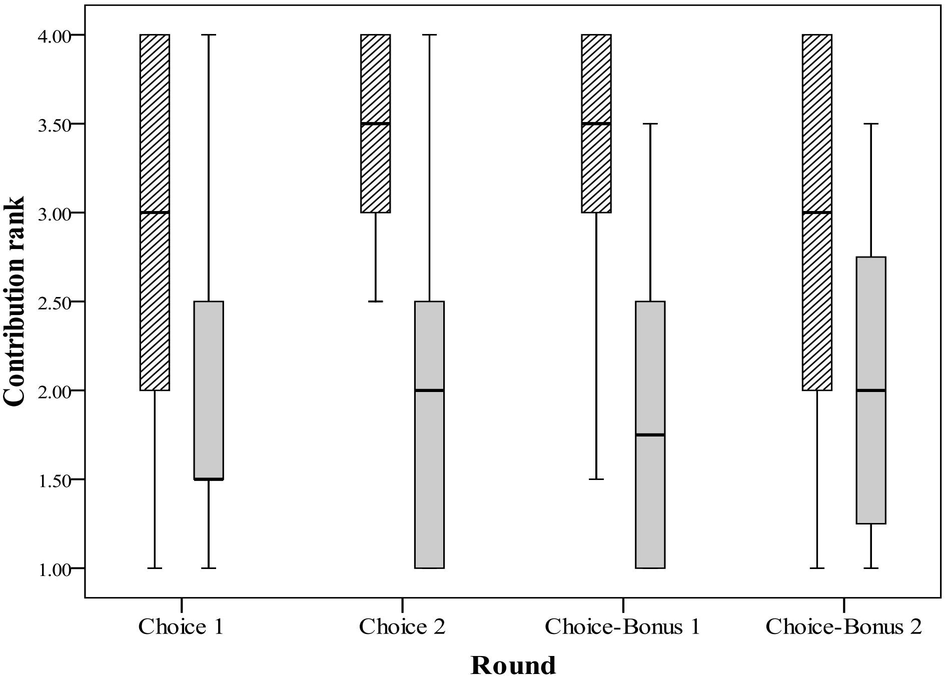


Figure 3.2 Boxplots presenting differences between contribution ranks of participants who played with assigned (striped boxes) or desired partners (grey boxes). Mann-Whitney U tests were significant at $p < 0.05$ in the Choice 1 round and at $p < 0.01$ in all other rounds.

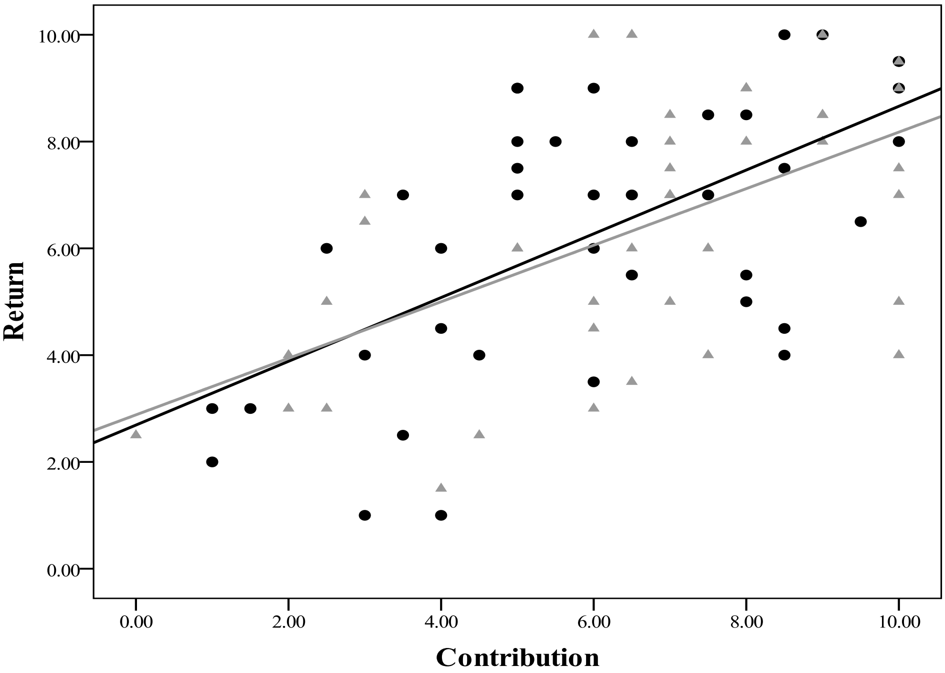


Figure 3.3 Relationship between participants' contributions in Stage 1 and their partners' contributions in Stage 2. Regression lines were fitted for the *Choice* condition (black line), Return = 2.69 + 0.59 * Contribution; and for the *Choice-Bonus* condition (grey line), Return = 2.88 + 0.57 * Contribution.

3.4. General discussion

The results of Study 1 support the main hypothesis that generosity is contingent on the incentive to build up cooperative reputation when the incentives are dictated by competitive altruism theory (Roberts, 1998). Cooperative contributions increased when they were going to be revealed to other players in comparison to when they remained anonymous (see also Barclay and Willer, 2007). This result alone is congruent with IR - individuals were able to reward others for their cooperation in Stage 2. However, when active choice of partners based on the information about others' contributions was possible, individuals displayed even higher cooperation levels. Hence, partner choice, the core ingredient of CA, invoked higher cooperation than a mere opportunity to interact in pairs with individuals of known reputations. Contrary to our expectations, there was no difference in contributions between the two conditions with partner choice but different reward levels, *Choice* and *Choice-Bonus*. We speculate that the payment method used in Study 1 did not allow for disentangling the two incentives. In a situation when even the slightest difference in profit can make someone a winner and leave others with nothing, what really matters is who outperforms others. Although participants could earn different number of lab pounds in the two conditions, more lab pounds in the *Choice-Bonus* condition did not translate to more money at the end of the game, as the highest earner would always receive £20. It is instructive to see how a manipulation of the payment method can affect participants' economic behaviour.

The results of Study 2 support the hypothesis derived from CA that those who develop a reputation for generosity acquire cooperative partners and receive more in return from them than less generous individuals (Roberts, 1998). To our knowledge this is the first empirical evidence for profits coming from reputation building in CA. In Barclay and Willer's (2007) study partner choice did not elicit high cooperation levels within interaction pairs and the authors suggested that repeated interactions were necessary to observe this effect: participants could otherwise build a cooperative reputation and then exploit it. In our study there was no incentive to exploit a partner in the *Choice* rounds, and cooperating was directly beneficial in the *Choice-Bonus* rounds. Nevertheless, a number of participants refrained from contributing. Even when the individual return from contributing to a public good exceeds the individual return from keeping the endowment, not all people behave in the optimal way (e.g. Saijo and Nakamura, 1995). Alternatively, despite detailed instructions and a trial round some participants might have not understood the rules of the game. According to research on individual differences in cooperativeness individuals belong to one of the three distinct groups that are stable over time: spiteful, payoff-maximizing and altruistic (Kurzban and Houser, 2005). Considering this, it is not surprising that those who were more cooperative in Stage 1 continued to be more cooperative in Stage 2, while less generous participants who could not play with their desired player acquired partners who refrained from contributing even though it was in their own interest.

As expected, reputational competition was stronger when it was possible for participants to receive a significantly higher payoff from partner choice; however, this effect was not significant when using a two-tailed test. We speculate that this may be because investing in a more costly signal is not always the best strategy (Bergstrom and Lachmann, 1997). Participants playing *Choice-Bonus* rounds might have decided that giving more would only increase the cost without providing greater long-term benefits. Alternatively, people may use different criteria for choosing partners than those proposed by CA. It seems logical to expect that the costlier the signal, the more attractive the sender of the signal is to the audience (90% of participants who filled in the questionnaire declared that they tended to choose individuals who contributed high amounts). However, one can well imagine that observers assess the value or the credibility of the signal in a different way for example by looking at the signal's consistency or its similarity to the observer's signals. Although in Stage 2 participants could see previous contributions from all Stage 1 rounds separately, they could remember the contributions and form a picture of how consistent other players were with regard to cooperative behaviour. In the post-study questionnaire two participants indicated they actually used this strategy when choosing partners. A different criterion used to select partners was choosing someone whose contributions were similar to contributions of the chooser which was a strategy used by one participant. Considering that two players could only make a pair if they chose each other and otherwise they were randomly assigned to a partner, this was a reasonable strategy likely to increase the chance of being paired with the chosen person. There is some evidence that positive assortative pairing occurs in the mating context where people pursue mates with a similar level of attractiveness (see e.g. Little et al., 2001). Such assortative pairing could also be present in the social context of acquiring partners for cooperative interactions .

In summary, we demonstrated that the level of generosity is dictated by the incentive to build up reputation predicted by competitive altruism theory. People do not make wasteful displays of generosity, that is, they refrain from cooperation when it does not serve any reputational purpose. Moreover, we showed that the short-term investments in reputation can be recouped in the long-term through the acquisition of desired partners and profitable interactions with them. Note that our experiment differs fundamentally from those on IR (e.g. Wedekind and Braithwaite, 2002, Milinski et al., 2002b) in that the benefits of reputation building come from assortative partner choice followed by directly reciprocal cooperative interactions. This is a crucial difference because it provides a different mechanism for reputation building. IR relies on the use of 'moral assessments' by which individuals decide who is a worthy recipient, despite never having the opportunity to receive back. CA, in contrast, relies on the benefits of obtaining the most profitable partnerships. Here, we empirically showed that people can indeed reap benefits from investing in reputation through CA. The significance of this finding is that displays of cooperation can be seen as an adaptive strategy, even when they are not reciprocated either directly or indirectly.

4. The effect of resource inequality on reputation-based cooperation in social dilemmas

Previous work shows that reputation-based competition for social partners can be a driving force for cooperation. This study examined how reputation-building behaviour varies according to differences in resources between individuals playing a social dilemma game. 60 students played a public goods game with three different endowments, knowing that their contributions would be revealed, and that they would have an opportunity to choose partners for a further game. Subsequently, participants made partner choices between two players whose endowments and contributions were displayed. Next, participants assessed players with different resources, who contributed different amounts to the common pool, with regard to their pro-social attributions and intelligence. We found that participants low in resources contributed proportionally more than wealthier participants. Further, those who devoted a larger proportion of their resources to others were judged to be more desirable social partners and were perceived as more pro-social but not more intelligent. This study demonstrates that within the context of reputation-based competition for partners, low-resource individuals invest relatively more in reputation than their high-resource counterparts and that in an environment with fluctuating resources partner choice is affected by the relative cost of an investment rather than its absolute value.

4.1. Introduction

While the literature on human economic behaviour shows that in apparently anonymous settings people behave in a moderately cooperative and fair way (Ledyard, 1995), the level of cooperation increases substantially when behaviour is public (e.g. Barclay, 2004, Milinski et al., 2002b). One possibility is that public cooperation can pay through developing a reputation which makes one more likely to be chosen for a profitable cooperative partnership, a process termed competitive altruism (Roberts, 1998). Such reputation-based competition for partners has been demonstrated in experimental studies (e.g. Barclay, 2004, Barclay and Willer, 2007). Most research on human cooperation employed paradigms in which all participants were endowed with the same amount of resources that they could spend to enhance private or common welfare. In studies where endowments were varied within a group of anonymous players, participants contributed to common welfare in proportion to the resources they were endowed with (for a review see Yu et al., 2009). Here we extended such studies by investigating, for the first time, how resource inequality affects cooperation levels in a setting where reputations may be important in partner choice. Furthermore, we explored a related question of how people estimate others' cooperative effort, pro-social attributions and intelligence when information about both the amount of possessed resources and the cooperative contribution is available.

4.1.1. Cooperation and fairness in social dilemmas

In the extensively modelled and used public goods game (PGG), players have an option to behave selfishly and keep monetary endowments to themselves, or to cooperate and contribute some of their resources to a common pool (Ledyard, 1995). The amount in the pool is multiplied by a factor higher than one and lower than the number of participants, and then distributed evenly among the players. If everyone contributes, all participants reap benefits. However, if some participants refrain from giving to the common pool or give little in comparison to others, they take advantage of those who contributed generously. Players' behaviour usually falls between the rational and the social optimum; i.e. the average contributions in the early stages of anonymous PGG range between 40% and 60% (Fehr and Schmidt, 1999). This is found even in anonymous games, where reputational incentives are excluded.

When social norms of cooperation and fairness are violated, group members are willing to impose costly punishment on the selfish individuals (Fehr and Gächter, 2002).

The social norm of cooperation also applies to a situation when participants have different opportunities to cooperate e.g. when due to the experimental setup they have different endowments. In this case, participants adjust their pro-social contributions to the amount of resources they possess. Participants with unequal resources tend to contribute a similar proportion of their endowments; consequently, in the absolute sense the input of financially privileged individuals to the public good is higher than the input of less privileged ones (Hofmeyr et al., 2007, van Dijk and Wilke, 1995, Yu et al., 2009, Wit et al., 1992, Van Dijk and Wilke, 1994). Such behaviour is socially desirable in the sense that participants with higher endowments are expected to follow the proportionality norm and contribute more than participants with lower endowments (Cress and Kimmerle, 2008).

4.1.2. The role of reputation in social dilemmas and partner choice

In an anonymous setting, participants with different endowments devote a similar proportion of their resources *pro bono publico* (Yu et al., 2009), but whether this holds true for situations in which people can build and use reputations has not been investigated. One theory as to why people should be concerned about their reputation is that of competitive altruism (CA; Roberts, 1998). This concept is based on the assumption that through costly advertisements of generosity, individuals increase their chances of forming successful cooperative partnerships with other individuals of high cooperative reputations. When individuals vary in cooperation, those who wish to interact with the most pro-social ones also need to acquire a cooperative reputation in order to be accepted as partners. Therefore, when playing economic games with reputational benefits in the form of future partnership formation, the level of cooperation is much higher than in anonymous games and non-anonymous games without the partner choice opportunity (Milinski et al., 2002b, Barclay, 2004, Barclay and Willer, 2007). It is known that cooperative individuals are indeed more often chosen as partners (Barclay and Willer, 2007) and, when the experimental design involves free pairing, they acquire the desired partners more often than the uncooperative ones (Sylwester and Roberts, 2010).

Only two studies have investigated how individuals with heterogeneous resources and different reputational investments were perceived as interaction partners. In Hardy and van Vugt's (2006) experiment participants watched four players (who were virtual and whose actions were pre-programmed) play PGGs with either high or low endowment. Participants were then asked to indicate the player who incurred the highest cost; to rate players with regard to their status; to choose the player they would like to interact with; and to distribute £5 between them and the other player. It was found that the player with low resources but relatively high contribution was perceived to incur the greatest cost; received the highest status; was most frequently chosen as a partner; and was allocated the largest fraction of £5 (Hardy and van Vugt, 2006). These results suggest that generosity is assessed in terms of the relative and not the absolute cost of the investment. However, the asymmetry of that study's design i.e. the fact that low-resource individuals could either contribute all of what they had or 20% of their resources whereas high-resource players could contribute either 10% or 50% of resources, did not allow for a situation in which high-resource players gave the same proportion of resources as low-resource players. The study left scope for exploring partner preferences when the choice occurs between players of different resources but identical proportions of cooperative contributions. In another study participants observed fictitious players punishing an uncooperative individual (Nelissen, 2008). Players who spent a high proportion of their resources on punishing were entrusted with more money than those who spent a low proportion, or than non-punishers, indicating that the relative cost of pro-social behaviour represents one's trustworthiness better than the absolute cost.

4.1.3. Perceptions of pro-social attributions and intelligence

When both the amount of resources and the contribution to the group welfare are known, perceptions of individuals can be shaped by two factors: their wealth and cooperation level. Although in an asymmetric PGG wealthier individuals tend to contribute absolutely more than the poorer ones (Yu et al., 2009), wealthiness in general is associated with *not* being likeable, kind, honest and willing to provide help (Christopher and Schlenker, 2000, Christopher et al., 2005). This suggests that poorer individuals, especially if they spend a large proportion of their resources, should be perceived in a more favourable light in terms of various pro-social attributions such as: trustworthiness, reliability, morality and cooperativeness. In contrast, wealthy individuals are generally perceived as more intelligent than poorer people (Christopher and Schlenker, 2000).

In psychology, uncooperative or exploitative behaviour has been traditionally described as Machiavellianism. People who score high on Machiavellianism are considered as more attractive and intelligent by colleagues than people who score low (Cherulnik et al., 1981). Interestingly, these perceptions do not reflect reality as Machiavellian intelligence is not correlated with general intelligence (Wilson, 1996). On the contrary, it has been proposed that altruism can function as a costly signal of general intelligence (Millet and Dewitte, 2007). Groups who scored higher on a reasoning test cooperated significantly more in Prisoner's Dilemma games than groups with lower scores (Jones, 2008).

4.1.4. Outline and hypotheses

Our experiment tested the predictions of CA in a scenario where players have unequal endowments. In such a setting, low-resource (LR) individuals are disadvantaged in the advertising stage of CA. In PGGs without reputation building, participants tend to contribute in proportion to their resources. However, to effectively compete for partners, the relative cost of a contribution for LR individuals needs to be much larger than for high-resource (HR) ones. It is important to note that this study did not seek to represent a situation in which differences between individuals' resources are large and permanent. Our study was designed to capture behaviour occurring among individuals of a similar class (e.g. students) who happened to possess different amounts of money (which does not mean that these amounts reflected their general wealth) and who were faced with a social dilemma (e.g. needed to collect money for some purpose).

We also investigated how participants' contribution decisions and resource levels affect their desirability as social partners. We examined preferences towards two players who contributed the same amount of money in the absolute sense (different relative cost of an investment) and preferences towards players who had different resources but contributed the same proportion of them (identical relative cost of an investment). Finally we tested how players with different resources and contributions are perceived by participants with regard to various pro-social characteristics and intelligence.

Hypothesis 1

In a CA setting (unlike in PGGs without reputational incentives in which individuals contribute in proportion to their resources) players with lower endowments will contribute relatively more than players with higher endowments. The contributions will reflect the cost required to reputationally compete with others. We predict that LR individuals will incur the highest relative cost and HR individuals the lowest relative cost.

Hypothesis 2

Preferences for social partners will be shaped by information about both their resources and their contributions. We predict that, when the absolute contributions are identical, participants will consider players who made a relatively costlier cooperative effort as more desirable social partners. The condition when contributed proportions of endowments are identical is exploratory; hence, we do not make any specific predictions.

Hypothesis 3

Judgments of pro-social attributions and intelligence will be affected by the relative rather than the absolute cost of cooperation. When two individuals contribute exactly the same amount of money but differ in their endowments, the one for whom the contribution was relatively costlier will be perceived as more cooperative, reliable, moral, and trustworthy than their counterpart. In contrast, the individual who spent a lower proportion of their resources on advertising generosity will be regarded as more intelligent. When both individuals spend the same proportion of their resources, there will be no judgement bias in terms of their pro-social attributions or intelligence.

4.2. Method

4.2.1. Participants and payments

20 groups of three (22 men: M age = 23.59, SD = 8.39 and 38 women M age = 23.37, SD = 5.93) were recruited for the study through e-mails and advertisements on the Newcastle University website. There were five male-male-female groups, six female-female-male groups, two male only groups and seven female only groups. Each person in a group was endowed with a different amount of experimental money, notionally £10 (LR), £15 (MR) or £20 (HR). Before the game started, participants were told that what they would earn in the game would be exchanged for real money at a fixed rate at the end of the experiment (participants were not told the exchange rate but were informed that an approximate average payment per participant in each group was £6). Because participants were arbitrarily provided with different amounts of experimental money, their final payoff would necessarily reflect those inequalities. In order to limit any disadvantage that players with lower endowment might feel, all participants were asked, after the experiment but before revealing the payoffs and debriefing, whether they wished to receive a payment according to their experimental earnings or a fixed amount of £6 for participation. Only 20% of participants chose the fixed payment. The mean payoff per participant was £6.29, SD = 0.88.

4.2.2. Procedure

On arrival, participants were seated at computers separated by partitions in order to ensure anonymity. Participants were informed that in Stage 1 they would play ten rounds of an economic game (which was in fact a PGG), after which their contributions would be revealed independently for each round to other players. It is important to note that participants were made to think that only their contributions (but not endowments) would be displayed. This created an incentive for LR individuals to invest in cooperation relatively more than HR individuals. In fact, both contributions and endowments were revealed before partner choice. Next, before the games were played, participants were told that their decisions from each round of Stage 1 would be accompanied by a request to choose a partner for one of ten rounds of another economic game played in pairs in Stage 2 (see Table 4.1). Participants were also told that only if two individuals chose each other, would they be allowed to play a round of the last game; if not, they would not play the Stage 2 round at all and would lose a chance to gain extra money. It was made clear that Stage 1 endowments would be independent of Stage 2 endowments. Such a design enabled us to represent a situation when resources fluctuate over time. Individuals were aware of the existing inequalities in resources, so e.g. a person endowed with £10 knew that the other two players had £15 and £20. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

Before the experiment started, participants were familiarized with the PGG and completed a short test in order to ensure that they understood the rules. Then, in Stage 1 of the experiment, participants played ten rounds of PGG, where they had an option to contribute any amount of money between 0 and the amount they were endowed with to a common pool or to keep this money to themselves. The amount in the common pool was doubled and shared equally among participants, so there was an opportunity to be selfish and free-ride on others' contributions, but there was also an incentive to contribute a lot in order to be chosen as a partner and play Stage 2 rounds.

Participants were then presented with ten successive choices to make between partners based on their endowments and PGG contributions. In order to investigate a broad range of combinations of endowments and contributions, we did not use the actual contributions made by other players in Stage 1 but instead substituted a predetermined series of values. In five presented choices participants could choose between two players who contributed identical absolute amounts (identical absolute contributions - IAC). In the other five rounds participants could choose between two players whose proportions of resources contributed to the pool were identical (identical proportional contributions - IPC). For example a person who was endowed with £15 in Stage 1, in one IAC round of partner choice could choose between Player A (contribution = 4, endowment = 10) and Player B (contribution = 4, endowment = 20). In contrast, in one IPC round this person could choose between Player A (contribution = 2, endowment = 10) and Player B (contribution = 4, endowment = 20).

Table 4.1 Description of the stages of the experiment distinguishing information given to participants before the study and the actual events participants experienced.

Stage	What participants were told about the stages of the experiment	What happened in practice
1 (PGG)	10 subsequent rounds of PGG without information about the contributions of other players.	10 subsequent rounds of PGG without information about the contributions of other players.
Partner choice	10 subsequent choices of partners for the 10 rounds of Stage 2 game played in pairs. For each choice, information about other players' contributions (but not endowments) of one round of the previously played PGG is presented.	10 subsequent choices of partners. For each choice, experimentally manipulated information about other players' contributions and endowments is presented in a balanced order.
2 (game in pairs)	10 rounds of an economic game played in pairs. Only if two players choose each other, are they allowed to play a round. If not, players skip a round and lose a chance of earning extra money.	Participants fill in a questionnaire. No Stage 2 game is played. Participants are debriefed and receive payments.

In the five rounds of the IAC condition, participants were presented with choices of LR, MR and HR players who contributed £2, £4, £6, £8, and £10, whereas in the 5 rounds of IPC participants were asked to make choices between players who contributed 20%, 40%, 60%, 80% and 100% of their resources. The choices were presented in a balanced order i.e. IAC rounds were alternated with IPC rounds. For example, a participant with a £10-allocation in Stage 1 made nine choice of partners between players with the following contributions/allocations: 2/15 vs. 2/20, 8/20 vs. 6/15, 8/15 vs. 8/20, 4/20 vs. 3/15, 10/15 vs. 10/20, 20/20 vs. 15/15, 6/15 vs. 6/20, 12/20 vs. 9/15 and 4/15 vs. 4/20.

After the partner choice stage, participants were given a short questionnaire to complete. In addition to demographic questions, participants were asked to decide who out of two fictitious individuals was more intelligent, reliable, trustworthy, moral and cooperative. Specifically, participants could select between Player A (endowment = 20, contribution = 4) and Player B (endowment = 10, contribution = 4), and then, between Player A (endowment = 20, contribution = 10) and Player B (endowment = 10, contribution = 5). Participants were also asked whether, after seeing Stage 1 contributions of others, they were satisfied with the amounts other players contributed (59.3% of participants were satisfied), and whether they expected such contributions or whether they were surprised with them (78% said they expected such contributions indicating that they considered them to be realistic).

Participants were informed that the Stage 2 game would not happen and their experimental profits would be calculated using the earnings from Stage 1, and that they might choose a fixed payment (£6) instead. No participant complained or wished to withdraw their data when the nature of the experiment was revealed to them.

4.2.3. Design and data analysis

The design was mixed in the sense that in the first part of the experiment three participants were endowed with different amounts of money (between-subjects factor) whereas in the second part all participants underwent the same experimental conditions in a balanced order (five choices in which the absolute cost was identical for the two presented players and five in which the relative cost was the same for them).

To analyze the relative cost of each player's contribution we calculated the proportion of their endowment that they contributed. We compared the proportions with ANOVA, using endowment as an independent factor. We ran the same analysis for the absolute contributions in order to investigate whether the absolute contributions directly reflected the endowments. The data from the partner choice part of the study and from the questionnaire were analysed using binomial tests that compared participants' choices to 50:50 distributions (no preference). Additionally, a relationship between Stage 1 endowment and partner preference was examined with chi square tests.

4.3. Results

4.3.1. Cooperation cost and endowment

As expected, we found a significant effect of endowment on the proportions of endowment contributed, *F*(2, 57) = 6.32, *p* = 0.003. Planned contrasts showed that having a higher than £10 endowment significantly decreased relative contributions, *t*(57) = -3.41, *p* = 0.001, but there was no difference between £15- and £20-individuals, *t*(57) = 1.019, *p* = 0.312 (see Figure 4.1).

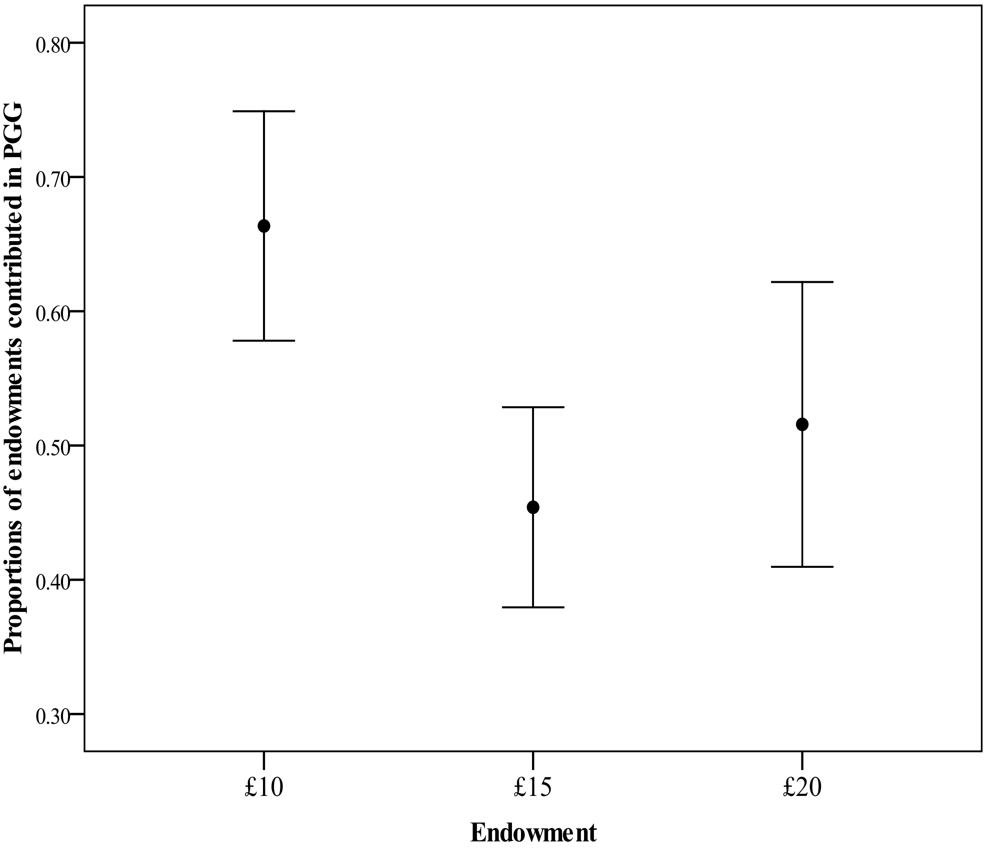


Figure 4.1 Mean proportion of endowment contributed in 10 rounds of PGG game by endowment level. The bars represent 95% confidence intervals.

We also found a significant effect of endowment on the absolute contributions, F(2, 57) = 8.74, p < 0.001. Planned contrasts revealed that while the absolute contributions of £20-individuals (M = 10.31, SD = 4.52) were significantly higher than those of £15-

individuals (M = 6.81, SD = 2.39), $t(28.81) = 3.06$, $p = 0.005$, the contributions of £10-individuals (M = 6.63, SD =1.82) and £15-individuals did not differ, $t(35.53) = 0.26$, $p = 0.796$.

4.3.2. Partner preference

On examining partner preferences in the IAC condition it is clear that participants tended to choose those who incurred a higher relative cost of cooperation (Table 4.2, last column).

Table 4.2 Frequency of selecting an individual who spent more on cooperation in the relative sense (IAC condition: lower-resource individual) and in the absolute sense (IPC condition: higher-resource individual) by participants endowed with £10, £15 and £20 in Stage 1.

	Amount					Total
	contributed	**£10**	**£15**	**£20**	$\chi2$	**chosen**
IAC	£2	95%	75%	65%	5.50	0.78**
	£4	90%	60%	70%	4.77	0.73**
	£6	90%	60%	60%	5.71	0.70**
	£8	90%	70%	75%	2.55	0.78**
	£10	90%	60%	70%	4.80	0.75**
IPC	20%	40%	65%	55%	2.55	0.53
	40%	40%	80%	60%	6.67*	0.60
	60%	50%	50%	85%	6.91*	0.62
	80%	50%	55%	85%	6.17*	0.63*
	100%	80%	70%	90%	2.50	0.80**

Percentages are displayed separately for different amounts/proportions contributed in the two experimental conditions. The penultimate column shows chi square statistics testing the association between one's level of resources and the desired partner's level of resources. The last column presents binomial tests comparing the proportion of participants who chose lower-resource individuals as partners in the IAC and higher-resource individuals in the IPC to the 50:50 distributions (no preference), * = $p < 0.05$; ** = $p < 0.01$.

For the IPC rounds, as the proportion of the contributed resources increased, participants were more likely to choose HR players; a clear preference is visible for rounds with 80% and 100% contributions (Table 4.2). In rounds where individuals offered 20%, 40% and 60% of their endowment, there was a trend towards choosing the HR player but no statistically significant preference. Allowing for multiple testing did not alter the significance of these results.

Using chi square tests we examined whether partner choices were affected by the Stage 1 endowment of £10-, £15- and £20-individuals (Table 4.2, penultimate column). For the five rounds of the IAC condition we did not find any significant effect, but there was a stronger tendency for £10-individuals than for £15- and £20-individuals to choose LR players. In the IPC condition, in the round with 40% contributions, 80% of £15-individuals preferred HR players. In the rounds with 60% and 80% contributions, 85% of £20-individuals chose HR players as preferred partners. In the round with 100% contributions, the majority of all three types of participants preferred HR players as partners but this trend was most pronounced in £20-individuals.

4.3.3. Trait assessment

When participants were asked to indicate the more intelligent player, in the IAC condition they chose the one who bore a relatively smaller cost of cooperation (Table 4.3). However, in terms of reliability, morality, cooperation and trustworthiness participants tended to select the player who bore a higher cost of cooperation. There was no preference for any player with regard to any characteristic mentioned above in the IPC condition. Interestingly, some participants were even reluctant to make the forced choice, explaining that both presented options were the same.

Table 4.3 Proportions of participants who chose the low-resource individual (LR) with regard to intelligence and pro-social attributions.

Characteristic assessed	LR chosen in the IAC	LR chosen in the IPC
Intelligence	0.23*	0.41
Trustworthiness	0.88*	0.52
Reliability	0.78*	0.50
Morality	0.85*	0.56
Cooperativeness	0.87*	0.53

Binomial tests compare the proportions to 50:50 distributions (no preference) * $p < 0.01$. In the IAC condition participants made a choice between individuals whose contributions/endowments were 4/20 and 4/10 whereas in the IPC condition the choice was between individuals with contributions/endowments: 10/20 and 5/10.

4.4. Discussion

Prior studies have empirically established that in anonymous interactions people with heterogeneous resources contribute a similar fraction of their endowment irrespective of the absolute value of this endowment (Yu et al., 2009). In this experiment we first investigated how different resources and the knowledge of the existing inequalities affect players' cooperative behaviour under the CA framework. The findings indicate that in the presence of reputational incentives LR individuals over-contribute to the common pool i.e. contribute more than a fair share. As expected, LR individuals contributed a significantly higher proportion of their endowments than MR and HR ones. If one had access only to information about participants' contributions but not endowments, it would not be possible to distinguish between LR and MR individuals as their absolute contributions did not differ. It would be difficult for LR individuals to compete with HR individuals who on average contributed £10.32, but LR individuals managed to effectively outperform MR. The fact that LR individuals spent such a large proportion of their resources on contributions to the common pool suggests that the relative cost of the cooperative investment depends on the perceived competitiveness of others. In the past, it has been shown that participants' behaviour in the PGG becomes more pro-social when reputational incentives in the form of partner choice exist (Barclay, 2004, Barclay and Willer, 2007). Here we found that when participants differ in available resources the poorest ones become the most generous in the presence of the incentive to be paired with the chosen partners.

An unanticipated result was that the relative cost of the contribution to the public pool did not differ between MR and HR individuals. Why did MR individuals not try to compete with HR when presented with an incentive to be able to play another economic game and earn extra money? It is plausible that there is a 'poorest in the group' effect and only the lowest-resource individual feels the pressure to devote more resources than others in order not to be seen as the least cooperative. Alternatively, LR individuals might have had the highest motivation to invest in reputation in Stage 1 and gain high profits in the paired game which could compensate for their earlier financial disadvantage.

Another important result is that participants paid attention to the relative cost of the cooperative investment, and showed a preference for players who gave a larger proportion of their endowment, supporting previous findings (Hardy and van Vugt, 2006, Nelissen, 2008). Preference for lower-resource high-contribution partners indicated that desirability is determined by the relative cost of cooperation. Such a partner preference can yield benefits in an environment where resources are prone to fluctuations because choosing poorer but more pro-social players over their richer but less pro-social counterparts will yield higher mutual benefits in the long run. An alternative motivation for people to choose the less wealthy but more pro-social players may be the willingness to punish HR individuals who gave no more than LR ones by preventing them from gaining the benefits from a cooperative partnership. This explanation seems reasonable considering people's willingness to spend money on punishing selfish individuals in a way to reduce resource-differences within a group (Fehr and Gächter, 2002).

In the condition where the absolute value of the contributions differed but the relative cost was the same, we found that the preference of higher-resource individuals depended on the cost of a cooperative act. There seems to be a conflict between preferring those who give a greater proportion of their endowment, on the one hand and preferring those who offer more in an absolute sense on the other (shown most clearly by the lower half of the last column in Table 4.2). In rounds where players contributed 20% to 60% of resources, both potential partners were considered equally attractive. This means that a greater absolute value of a contribution is not preferred even when it represents the same proportion of different endowments. There is, therefore, a bias favouring LR players. However, as the differences in the absolute value of contributions between potential partners increase, there comes a point (80%) where the absolute contribution becomes the choice criterion. Economic game experiments have shown that the commonly accepted as fair and most frequently given contribution in early rounds of PGG is 40%-60% of one's resources (Fehr and Schmidt, 1999, Ledyard, 1995). Interestingly, the shift in partner preference occurs at a point at which both players should be regarded as unexpectedly generous. Thus, in the case of equal proportions of contributions, the absolute cost of a contribution appears to affect choice decisions only when both low- and high-resource individuals are extremely generous i.e. they are contributing much more than a fair share.

We found that the amount of resources one was allocated in Stage 1 did affect partner choice decisions in the IAC condition to a certain extent; yet none of the chi square tests was significant. A noticeable but non-significant bias in £10-participants to choose MR individuals (90-95% frequency in all IAC cases) suggests that participants who experienced competition with wealthier individuals paid more attention to the contribution/endowment ratio than others and appreciated the effort of less wealthy individuals sacrificing a high proportion of their resources for the collective good. In terms of the choice between proportionally identical contributions, there was a bias in £20-

individuals to choose MR individuals. It appears that although participants had known that endowments in Stage 1 would be reassigned in Stage 2, they favoured individuals of more similar resources. In-group bias predicts that people will favour individuals of their own social group (see Yamagishi and Mifune, 2009). In this study participants could not select individuals of the same endowment class but participants' choices reflect a willingness to minimize the distance between theirs and their partner's endowment.

As hypothesized, the relative and not the absolute cost of the investment determined which player was perceived as more cooperative, trustworthy, reliable and moral, supporting the earlier findings (Christopher et al., 2005, Nelissen, 2008). In line with our predictions and reports on the perceptions of people with Machiavellian skills (Wilson et al., 1996), HR individuals who made less costly investments and kept more resources to themselves were considered as more intelligent than their LR counterparts. Interestingly, when the proportions of contributions did not differ between individuals the perceptions of their intelligence and pro-social attributions were similar. Such perceptions indicate that when assessing pro-social attributions and intelligence people base their judgements on the relative and not the absolute cooperative effort, but note that these perceptions might vary depending on the way in which people acquired the resources (Smith et al., 2003b).

One of the most significant findings to emerge from this study is that individuals with low resources, when set in a reputation building context, try to keep pace with more resourceful players by displaying generosity that is relatively more costly for them than for others. These individuals appear to foresee the long-term benefits coming from reputation, and decide it is best to invest much more in it in order to be able to compete with others (Roberts, 1998). Such a strategy can pay off, because later, the relatively high-contributing LR individuals would be preferred for dyadic interactions and would be able to acquire the most cooperative partners (Sylwester & Roberts, 2010). The evidence from this study suggests that people take into account the relative cost of an advertisement as well as its absolute value. Moreover, the relative cost of the investment is positively associated with being perceived as cooperative, trustworthy, moral and reliable, but not intelligent. In future work, using a wider range of endowments could help to establish whether the effect of increasing the relative cost of an investment applies exclusively to the poorest individuals.

5. The effectiveness of competitive altruism and indirect reciprocity at sustaining cooperation in social dilemmas

Cooperation in social dilemmas can be maintained through a reputational mechanism of indirect reciprocity (IR) – a cooperator is likely to become a recipient of cooperative acts from others. Here, we propose another mechanism which aids in re-establishing group cooperation after a decline: competitive altruism (CA), which proposes that cooperators benefit through access to cooperative partners and long-term payoffs from direct interactions with these partners. 20 groups of four students played a series of public good games which were (1) not alternated with any other game, (2) alternated with an indirect reciprocity game or (3) alternated with a direct reciprocity game preceded by partner choice. The length of the indirectly and directly reciprocal dyadic interactions was varied. We found that contributions to the common pool were higher in the CA than in the IR setting. When group games preceded long-term rather than short-term dyadic interactions this effect was even more pronounced. Individuals received more from partners in direct than indirect games which can explain strong investment in reputation before direct interactions. Our findings indicate that strategic reputation building through CA can be a more robust mechanism for maintaining cooperation than IR.

5.1. Introduction

Social dilemmas occur when collective interests are in conflict with private interests. Numerous studies have shown that, in an anonymous experimental setting, people presented with social dilemmas initially tend to cooperate but over time cooperation declines (e.g. Ledyard, 1995). In contrast, repeated real life voluntary contributions such as charity donations or volunteering (Penner et al., 2005) remain at a relatively stable level. What mechanism causes people to continue behaving in a pro-social way?

From an economic perspective, the most rational choice for an individual is the one resulting in them receiving the maximum possible profit. Such a choice usually has a negative impact on the profits of others which causes a social dilemma. When, instead of investing in public resources, individuals have an opportunity to use them, a 'tragedy of the commons' can occur with people overexploiting the common goods (Hardin, 1968). To explain how cooperation in groups could be maintained when individuals are faced with a social dilemma two key mechanisms have been proposed: altruistic punishment (Fehr and Gächter, 2002) and indirect reciprocity (Milinski et al., 2002b). Here we consider another way in which cooperation could be enhanced: reputation-based partner choice, also described as competitive altruism (Roberts, 1998).

Social dilemmas have been extensively modelled with the use of a public goods game (PGG). In a standard design, four individuals receive an endowment from the experimenter. They can either keep the whole endowment or contribute all or some part of it to the common pool. The amount collected in the pool is multiplied by a factor larger than one and smaller than the number of participants. Then, the amount is shared equally among all individuals irrespective of whether they have contributed (Ledyard, 1995). In PGG each individual is tempted to keep their endowment and enjoy the benefits of others' contributions to the pool. However, the group would be best off if everyone contributed their whole endowment. When playing PGGs people initially contribute between 40% - 60% of their endowments (Ledyard, 1995). After a few rounds contributions tend to decline often resulting in everyone keeping their endowment. Research has shown that cooperation in PGG can be maintained when individuals can punish or reward others for their contributions (Rand et al., 2009, Fehr and Gächter, 2002).

Another way in which cooperation could be re-established after a decline is due to participants' motivation to invest in a cooperative reputation. Individuals may be encouraged to build up cooperative reputation in order that others indirectly reciprocate their cooperative acts. Indirect reciprocity theory (IR) proposes that altruistic acts are reciprocated not by the recipients of altruism but by others who have access to a donor's cooperative history (Alexander, 1987). In recent publications the definition of IR is narrowed down and it is assumed that in IR individuals with high cooperative reputations are 'rewarded' with cooperation from others who by cooperating with cooperators increase their own reputation. IR, therefore, predicts that cooperation is directed to cooperative individuals (see e.g. Milinski et al., 2002b).

Experimental studies have provided evidence that people do indeed reward those who are generous to others (Seinen and Schram, 2006, Milinski et al., 2001, Wedekind and Braithwaite, 2002). Milinski et al. (2002b) showed how, thanks to reputation through IR, initial levels of cooperation can be successfully rebuilt following a decline. In their experiment, participants either played PGGs which were alternated with IR games during the whole experiment or, after playing eight consecutive rounds of PGGs participants played iterated IR games. In the former case the level of cooperation remained high throughout the 16 alternated rounds whereas in the latter case, when, after eight PGG rounds, IR games were introduced, cooperation increased to the initial high level (Milinski et al., 2002b).

An alternative to the IR mechanism through which individuals can reap benefits from investing in reputation is competitive altruism (CA; Roberts, 1998). According to CA theory people build up reputations in order to be chosen by other cooperative individuals for profitable partnerships (Sylwester and Roberts, 2010). In CA individuals try to outperform each other in cooperation so that they acquire the best partners for dyadic interactions. Cooperative acts can be directed to individuals of low or no cooperative reputation, it is only important that cooperative and desirable social partners observe such displays. Hence, unlike the current interpretation of IR (see Chapter 2), CA does not assume that recipients of cooperative acts have themselves high cooperative reputation. In CA cooperative investments may function as costly displays or handicaps *sensu* Zahavi (1997). The more one spends on cooperative displays (even towards individuals of low or unknown cooperative reputations), the more attractive a social partner one makes (Barclay, 2004, Barclay and Willer, 2007). Considering this and the fact that the benefits from long-term interactions with a cooperative partner are likely to be a higher incentive to invest in reputation than one-off benefits from indirectly reciprocated acts we tested the idea that CA works better than IR at re-establishing cooperation in a social dilemma situation.

5.2. Method

20 groups of four students (35 male, mean age = 24.11, SD = 5.37 and 45 female, mean age = 21.84, SD = 4.95) were recruited from Newcastle University to participate in a decision-making study with monetary rewards. Participants were taken one by one to the computer lab and accommodated in cubicles so that they were anonymous to each other. A questionnaire revealed that 88.8% of participants did not have any visual contact with other players. Participants were told that they would have £300 lab pounds to play with for the whole experiment, and that what they earn would be exchanged to real money at the rate £1 real pound = £50 lab pounds at the end of the study. In each round participants could use up to £10 lab pounds. On average participants earned £6.73. The experiment was conducted using z-Tree software for economic games (Fischbacher, 2007).

Before the experiment started participants read the instructions and took a quiz designed to test whether they had understood them. Specifically, participants were asked to calculate outcomes of the games for fictitious players who behaved either selfishly or cooperatively (see Appendix B). During the experiment, instead of real names participants were known by neutral nicknames (names of chemical elements). After each round all participants' decisions and profits for the most recent round were displayed publicly for 30s. The three games used in the study, public goods game (PGG), direct reciprocity (DR) and indirect reciprocity (IR) are described below:

PGG (referred to as 'Group game' in the study)

Participants had an option to contribute £0-£10 to the common pool. The amount in the pool was multiplied by 1.5 and shared equally among 4 players.

DR (referred to as 'Two-way game' in the study)

Participants could choose a partner for this game. If two players chose each other they would be informed that their desired partner chose them as well and they would play together. If the desired player chose someone else, participants were informed that the person they chose did not wish to interact with them and that a partner would be assigned to them arbitrarily. Participants were told the nickname of the player they were assigned to. When in pairs participants had an option to simultaneously give £0-£10 to their partner. The amount given would be multiplied by 1.5 before it reached the recipient.

IR (referred to as 'One-way game' in the study)

Participants could give £0-£10 to an arbitrarily indicated player. The amount given was multiplied by 1.5 before it reached the recipient. Participants knew that there would be no direct reciprocity in this game i.e. if Potassium was a potential donor to Carbon, Carbon would never be a potential donor to Potassium but another participant would be able to give to Potassium.

All twenty groups played five rounds of PGG at the beginning and five rounds of PGG at the end of the experiment. What differed between the two conditions were the middle games which either reflected the indirect reciprocity (IR) or competitive altruism setting (DR). Ten groups played five rounds of PGG with each round followed by a DR round and five rounds of PGG with each round followed by an IR round. The other ten groups played two rounds of PGG, each followed by four rounds of DR with the chosen/assigned partner and two rounds of PGG, each followed by four rounds of IR. In summary, in a half of the groups reputation gained in a PGG round was linked to a single DR or IR round (short-term dyad condition) whereas in the other half it was linked to four DR or IR rounds (long-term dyad condition). Additionally, to control for order effects, one half of the groups played the competitive altruism condition (PGG alternated with DR) first whereas the other half played the indirect reciprocity condition first (PGG alternated with IR). We expected a general drop in contributions over time; hence, any alternation that occurred later in the experiment would have a weaker impact on maintaining high cooperation. In total, participants played 30 rounds of different games (see Table 5.1). The details of the procedure can be found in Appendix B.

Table 5.1 Examples of four possible sequences of playing the games.

five initial rounds					PGG rounds alternated with IR or DR																				five final rounds				
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
□	□	□	□	□	□	○	□	○	□	○	□	○	□	○	□	△	□	△	□	△	□	△	□	△	□	□	□	□	□
□	□	□	□	□	□	△	□	△	□	△	□	△	□	△	□	○	□	○	□	○	□	○	□	○	□	□	□	□	□
□	□	□	□	□	□	○	○	○	○	□	○	○	○	○	□	△	△	△	△	□	△	△	△	△	□	□	□	□	□
□	□	□	□	□	□	△	△	△	△	□	△	△	△	△	□	○	○	○	○	□	○	○	○	○	□	□	□	□	□

□ = PGG, ○ = DR, △ = IR

Our design was close to Milinski et al.'s (2002b) in that we asked groups of participants to play economic games with each other and we alternated PGG rounds with IR rounds. Additionally, we alternated PGG rounds with direct reciprocity (DR) rounds for which participants could choose a desired partner. We also investigated a situation when partner choice led to a longer-term relationship by introducing a few DR rounds played with one chosen partner. The main methodological difference between ours and Milinski et al.'s setting was that in our games participants could contribute or donate any amount between £0 and £10 while Milinski et al. used a discrete contribution method – contribute all or nothing. Continuous donations better represent attempts to collect money in the real world because the available resources are often divisible. Theoretical studies have shown that cooperation is more likely to occur when individuals can calibrate their cooperative investments instead of making all-or-nothing decisions (Roberts and Sherratt, 1998, Killingback et al., 1999, Roberts and Renwick, 2003). Participants in our study could see others' contributions but we also provided them with information about how much other people earned in each round which was likely to increase within-group competition (Nikiforakis, 2010).

5.3. Results

5.3.1. Main analysis

For the main analysis we used each group of four participants as a statistical unit. First, in order to show a decline in contributions over time we compared mean contributions in the first and the fifth round of PGG with the paired samples t test. Contributions were significantly higher in the first round (M = 5.45, SD = 1.58) than in the fifth round (M = 2.95, SD = 2.45), t(19) = 5.93, $p < 0.01$. Similarly, we found that, when considering the final PGG rounds, contributions in the first round (Med = 2.37, IQR = 3.0) were higher than in the last round (Med = 1.12, IQR = 2.19), z = -2.21, $p < 0.05$ (due to the lack of normality in the distribution of the variables a Wilcoxon test was used). Contributions in the first round of a PGG are usually high because of participants' initial lack of experience or 'confusion' (see Andreoni, 1995). The decline in contributions over time can be explained by conditional cooperation i.e. fine-tuning the level of cooperation to others (Fischbacher et al., 2001). For this reason we calculated mean contributions in the initial five and the final five rounds and treated them as reference levels. The analysis involved mean PGG contributions in four different contexts: the first five rounds, the rounds alternated with DR, the rounds alternated with IR and the last five rounds. A mixed 2(order: DR or IR first) × 4(context) ANOVA applied to groups who were assigned to the short-term dyad condition showed a significant effect of context on PGG contributions, F(3, 24) = 4.29, p = 0.015. In order to investigate which of the two contexts, IR or CA enhanced cooperation in PGG more we did planned comparisons. We used a Bonferroni corrected significance of $\alpha^* = \alpha/N$ where N = number of comparisons (hence $\alpha^* = 0.05/3 = 0.016$). In the short-term dyad condition, contrasts revealed that contributions were significantly larger when PGG was alternated with DR rounds than with IR rounds, F(1,8)

= 13.19, p = 0.007 (see Figure 5.1). Moreover, in the competitive altruism context contributions were on average as high as in the first five rounds, F(1,8) = 3.06, p = 0.118 while when PGG rounds were alternated with IR, contributions were significantly lower than in the first five rounds, F(1,8) = 22.58, p = 0.001.

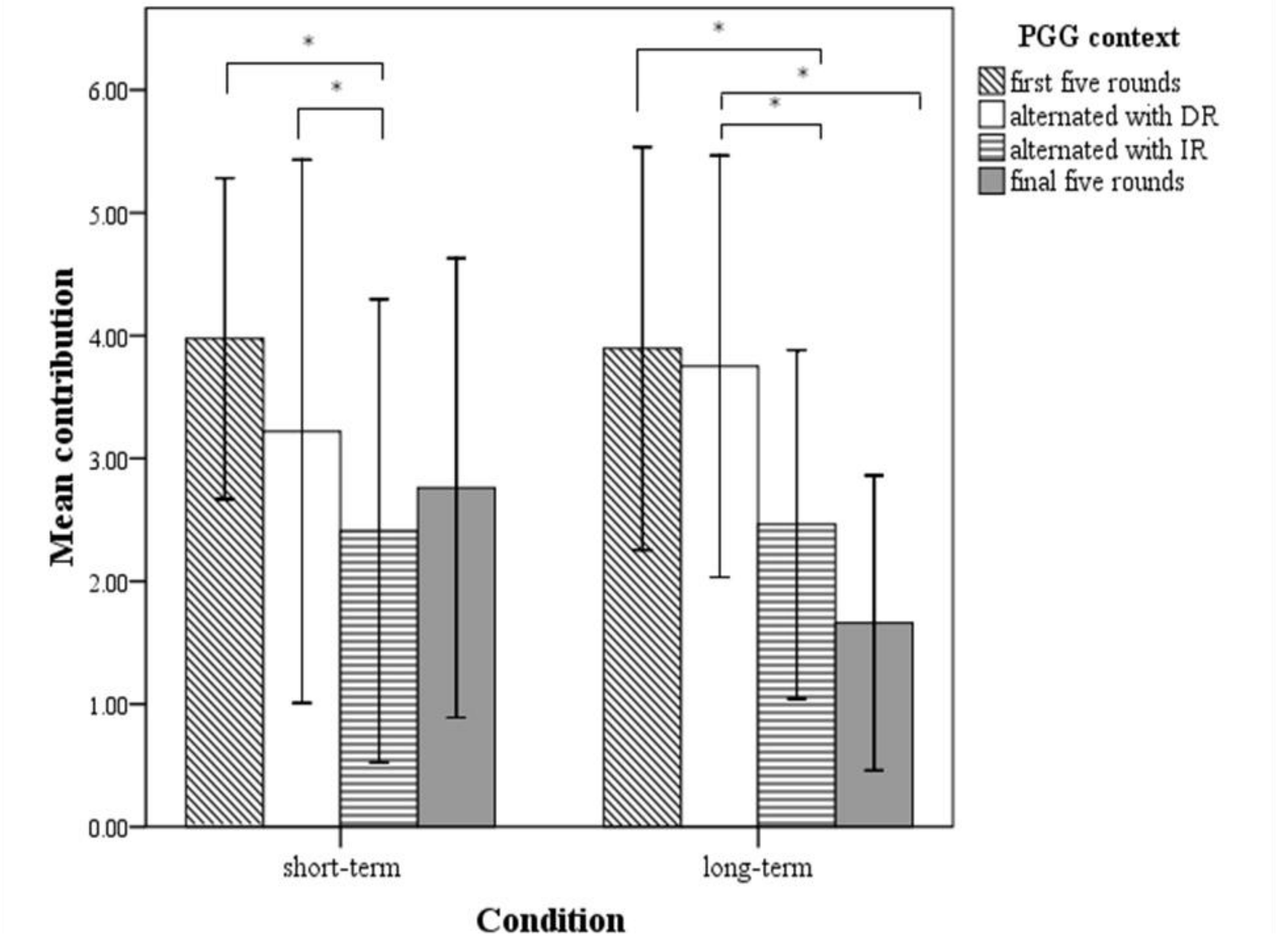


Figure 5.1 Mean contributions to PGG by game context in short- and long-term dyad condition. Bars represent 95% confidence intervals, * p < 0.016.

There was also a significant interaction between the context of playing PGG and the order of alternating with DR and IR rounds F(3, 24) = 5.40, p = 0.006. This indicates that contributions in the four investigated contexts were affected by whether PGG was first alternated with DR or IR rounds. The difference in contributions between PGG alternated with DR and IR was more pronounced in groups where the alternation with DR happened before the alternation with IR, F(1, 8) = 26.59, p = 0.001 (see Figure 5.2).

A similar 2×4 mixed ANOVA applied to groups in the long-term dyad condition showed a significant effect of context on PGG contributions, F(3, 24) = 7.86, p = 0.001. Contrasts revealed that contributions were significantly larger when PGG was alternated with DR rounds than with IR rounds, F(1,8) = 10.95, p = 0.011 (Figure 5.1). Contributions in PGG alternated with DR did not differ from contributions in the first five rounds, F(1,8) = 0.08, p = 0.79, whereas contributions in PGG alternated with IR did, F(1,80 = 11.54, p = 0.009. In the long-term dyad condition there was no interaction between the order of playing PGG alternated with DR vs. IR and contributions to PGG in different contexts, F(3, 24) = 1.50, p = 0.241.

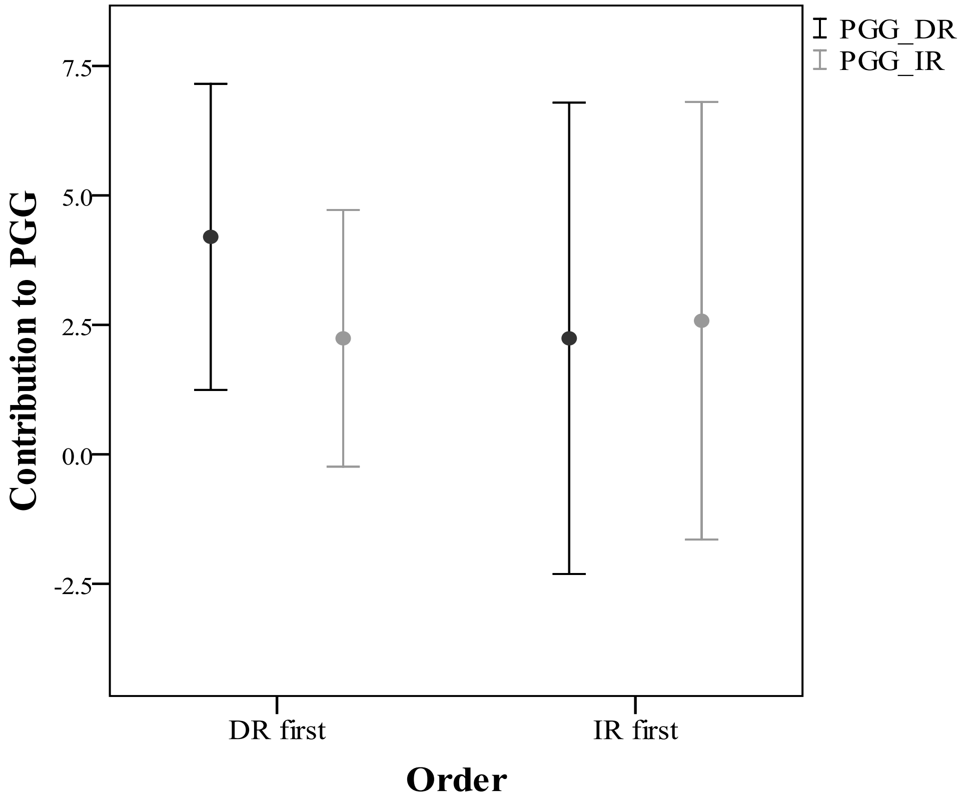


Figure 5.2 Means and 95% confidence intervals of contributions to PGG followed by DR or IR in groups where alternation with DR or IR happened first

The short-term dyad condition involved five choices of partners. In order to investigate whether the choices were consistent across rounds we conducted four chi square tests comparing the observed proportion of participants who chose the same or a different player as in the previous round and the expected proportion. Because each participant could choose among three potential partners we expected that one third of participants would choose the same partner by chance. All tests were significant showing a high consistency in participant's choices (see Table 5.2). Finally, we investigated whether experience affected partner choice, that is, whether participants were likely to choose a partner they played with in the previous round. In all four cases half of participants wished to interact with a participant they played with in the previous round which was significantly more than expected (Table 5.2). We also examined the consistency in choices and the effect of experience in the long-term dyad condition which involved only two choices each followed by a DR game. The proportion of same choices was not significantly different from the expected values (observed: 18, expected: 13.3), $\chi^2 = 2.45$, $p > 0.05$. Participants did not choose the partner they played with in the previous four rounds of DR with a probability higher than chance (observed: 17, expected: 13.3), $\chi^2 = 1.51$, $p > 0.05$.

Table 5.2 A comparison of observed and expected choices of the same partner that was chosen in the previous round and choices of the same partner with whom one played with in the previous round in DR games (short-term dyad condition).

comparison between rounds	observed number of participants who made the same choice of partner in two consecutive rounds	observed number of participants who chose the partner they played with in the previous round	expected values	Chi square for interaction between two consecutive choices	Chi square for interaction between previous game partner and current partner choice
1st vs. 2nd	26	20	13.3	18.05**	5.00*
2nd vs. 3rd	23	20	13.3	10.51**	5.00*
3rd vs. 4th	22	20	13.3	8.45**	5.00*
4th vs. 5th	22	20	13.3	8.45**	5.00*

* $p< 0.05$, **$p<0.01$

Table 5.3 A comparison of donations received in DR game between participants who decided to stay with a partner they played with in the previous round and those who preferred to change the partner in the short-term dyad condition.

Comparison between rounds	T	M(SD) stayed with partner	M(SD)changed partner
1st vs. 2nd	-1.55	6.0(3.45)	4.40(3.08)
2nd vs. 3rd	-1.20	5.30(3.18)	3.95(3.89)
3rd vs. 4th	-3.05**	6.1(3.45)	2.8(3.40)
4th vs. 5th	-2.15*	5.7(4.05)	3.2(3.25)

* $p< 0.05$, **$p<0.01$

In the traditional one-shot design of competitive altruism individuals build reputation in one stage and use it in the next one. In this experiment, because of repeated DR rounds, participants could both use a reputation built earlier in order to acquire a cooperative partner and by behaving cooperatively, build reputation for the future rounds. Hence, partner choice could have been determined by reputation gained in DR rounds.

Indeed, we found evidence suggesting that high donations from a partner in DR games contributed to this partner being chosen again. In the short-term dyad condition, participants who had received higher donations from partners in DR tended to choose the same partners again while those whose partners had given lower donations tended to choose different players for the next round (a trend present in the first two comparisons and a significant difference in the last two in the short-term dyad condition, see Table 5.3). Similarly, in the long-term dyad condition we found that participants who received on average higher donations from a partner they had played with in the previous four rounds (M =5.82, SD = 2.98) tended to choose this partner again while players who had received on average lower donations (M = 3.73, SD = 2.52) were inclined to choose a different partner for the next four rounds, t(38) = -2.40, p = 0.021.

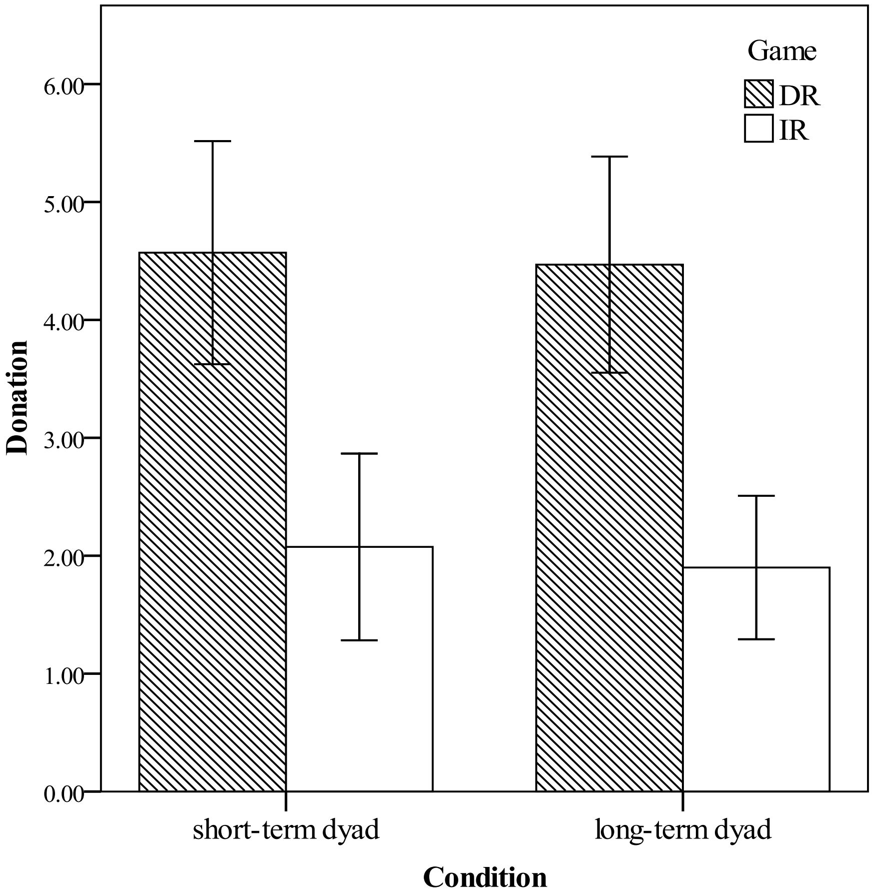


Figure 5.3 Mean donations to a partner in DR and IR rounds by condition. Bars represent 95% confidence intervals.

We compared the mean amounts given to partners in DR and IR rounds irrespective of whether in DR individuals were paired with their desired partners. A paired samples t test indicated that in the short-term dyad condition, donations in IR (M = 2.07, SD = 2.46) were significantly lower than donations in DR (M = 4.57, SD = 2.42), t(9)= -4.45, p <0.01, (log transformation was used to satisfy the assumption of normality). A similar effect was found in the long-term dyad condition: DR (M = 4.47, SD = 2.13) vs. IR (M = 1.90, SD = 1.66), t(9)= 4.63, p = 0.001, (see Figure 5.3).

5.3.2. Additional analysis

It was not the aim of our study to empirically test CA theory but rather to assess its relative power in comparison to IR in terms of enhancing cooperation in social dilemmas. Nevertheless, we did a few additional analyses to test whether the predictions of CA held as in previous studies. We focused on the first round of PGG alternated with partner choice followed by direct reciprocity (DR) because partner choices in the further rounds were likely to be affected by reputation built not only in the PGG but also in the previously played DR rounds.

We first tested whether the most cooperative players in the PGG were more likely to be chosen as partners for the following DR game. PGG contributions in each group were ranked from 1 (lowest contributor) to 4 (highest contributor). We found a significant positive correlation between contribution rank and the number of times a player was chosen as the desired partner by others in the group, r_S (38) = 0.31, p < 0.05 (short-term dyad condition), r_S(38) = 0.61, p < 0.01(long-term dyad condition). We expected that the tendency to rely on costly investments in reputation when choosing partners would be more pronounced in the long-term than in the short-term dyad condition. This is because we inferred that the advantage of choosing a cooperative partner for four rather than one round of a direct exchange would be higher; therefore the motivation to profit from the direct exchange would be reflected in the strength of correlation. By calculating the difference between the two correlation coefficients using the Fisher r-to-z, we found a non-significant tendency for the correlation in the long-term dyad condition to be stronger than in the short-term dyad condition, z = -1.67, p = 0.09. We also investigated whether the most cooperative individuals were more likely to be paired with their desired partners. The ranked contributions were recoded in two groups: ranks 1, 1.5, 2 and 2.5 were categorized as low contributors, ranks 3, 3.5 and 4 as high contributors. Due to the low expected count we performed one chi square test on data from both short- and long-term dyad conditions. We found a significant interaction between players' contribution category and being accepted by a desired partner, χ^2 = 6.88, p < 0.01 (see Table 5.4 for exact values).

Table 5.4 Cross-tabulation showing the number of participants with different contributions who were accepted or not by their desired partners.

	Played with the desired partner	Did not play with the desired partner
Low contribution	14	35
High contribution	18	13

Finally, we were interested in whether the opportunity to play with the chosen partner for four rounds in contrast to one round would provide a higher reputational incentive and elicit higher cooperation. Contributions in the first PGG round in the CA context in both short-term and long-term dyad conditions (round number 6 in Table 5.1) were aggregated by group. Participants tended to contribute more in the long-term (M = 4.1, SD = 2.67) than in the short-term (M = 3.42, SD = 3.09) condition, however, this trend was not significant, t = -0.52, > 0.05. A larger sample size might be needed to capture this effect.

5.4. Discussion

Our results show that the opportunity to choose partners for a dyadic interaction is a stronger incentive to invest in reputation than IR. This finding suggests that CA is a more efficient mechanism than IR for maintaining high contributions in social dilemmas, and possibly for explaining the evolution of reputation-based cooperation. In the present study we only compared strategic IR with CA. However, it has been shown that IR exists even without reputational incentives in which case it is called 'pure IR' – helping in order that the recipient helps another individual while the helper's image score is not made public (Engelmann and Fischbacher, 2009). Considering that cooperation levels in pure IR are lower than in strategic IR (Engelmann and Fischbacher, 2009); and that humans did not evolve in an environment where individuals' cooperative reputations were kept private, the mechanisms involving strategic reputation building seem to provide a more plausible explanation for the development of human cooperation than those involved in IR. Unlike IR, CA is purely strategic in assuming that those who cooperate more will acquire more cooperative partners and form more beneficial partnerships than selfish individuals (Roberts, 1998).

In our experiment, IR was clearly less efficient at re-establishing high levels of cooperation than in Milinski et al.'s (2002b) study. In previous research, the continuous contribution method we used here was found to induce more cooperation in PGG with a provision point (where participants' contributions need to exceed a certain amount in order for the group to receive a reward) than the discrete all-or-none contributions (Cadsby and Maynes, 1999). It is not then likely that our contribution method lowered contributions in PGG when it was followed by IR. On the other hand, we not only presented all participants with information about contributions of all players but also their profits in each round which might have led to reduced levels of cooperation in general (Nikiforakis, 2010), but again it should not have affected exclusively the PGG followed by IR. Milinski et al. (2002b) showed that (1) when social dilemmas are alternated with IR from the very beginning, cooperation levels in PGG remain stable; (2) while cooperation in iterated social dilemmas drops, when iterated IR games are introduced, it increases and remains at a high level in IR (*not in* PGG, hence PGG is not a reference level). In contrast, our study investigated whether, after a decrease in contributions in iterated social dilemmas, alternating them with IR or DR rounds can restore cooperation (the reference level was always cooperation in PGG and not in IR or DR games). It is then worth stressing that the question of whether cooperation in social dilemmas can be rebuilt after a decline can only be answered using our design. This results from the fact that in Milinski et al.'s (2002b) design the decline in cooperation in PGG rounds was followed by an increase in cooperation in IR rounds and not in PGG rounds alternated with IR.

Interestingly, in the condition in which individuals formed short-term one-round dyads in DR and IR, the order in which PGG was alternated with DR and IR strongly affected PGG contributions. When, after the initial five PGG rounds, participants played PGG rounds alternated with IR rather than DR, the alternation with IR yielded similar levels of cooperation to the alternation with DR. One possible explanation for this pattern is that cooperation dropped so much after five PGG rounds and five PGG rounds alternated with IR rounds, that short-term dyadic interactions in CA were too weak an incentive to restore it. Because individuals known by nicknames started acquiring reputation from the first round, cooperative decisions and partner choices were based not only on the behaviour in the preceding PGG round but also on the overall cooperative image built throughout the game.

In the short-term condition participants were consistent in their choices of partners for DR rounds and often played with the partner from the previous round. Such a consistent matching system made the short-term dyad condition similar to the long-term dyad condition in the sense that the likelihood of staying with the same partner for a few rounds, despite updating the choices in each round, was high. Interestingly, in the long-term condition the two choices made were not consistent. Partner choice was clearly affected not only by reputation built in PGG but also by experience and the possibility of partner control (encouraging cooperation with a threat of terminating an interaction) in repeated DR rounds: participants who received relatively high donations from their partners in DR rounds were likely to stick with those partners and choose them again. In long-term CA, individuals can use both reputation and experience to decide who would make a good partner.

Investing in reputation in the PGG rounds preceding DR rather than IR can be explained by higher donations to partners in DR than in IR games regardless whether participants were allowed to play with the desired partner or not. Participants preferred to give to someone who had a chance to reciprocate directly rather than indirectly which suggests that they were more willing to build first-hand rather than second-hand reputation. This result supports previous models which showed that when a possibility of re-meeting exists individuals are more likely to rely on experience than reputation (Roberts, 2008). Therefore, it pays off to invest more in reputation when there is an opportunity for direct interactions.

In conclusion, this study extended research by Milinski et al. (2002b) by showing that the mechanism of CA allows for restoring initial cooperation levels better than IR. Participants invested more in reputation used for DR than IR games and received more cooperation in DR than IR games. All effects appear to be stronger when there is a possibility of establishing long-term partnerships. Human CA has been shaped in a natural environment where a chance of re-meeting was high, long-term social interactions could occur, and because of time constraints only a few out of many potential partners were needed (e.g. the number of people with whom an individuals can maintain a stable social relationship is limited, see Dunbar, 1998) . This study showed for the first time that the opportunity to form long-term dyadic interactions with a desired partner can have a considerable effect on sustaining cooperation in social dilemmas.

6. Memory bias for reputational gossip: Cheater-detection or rarity-detection?

Language enables indirect monitoring of others' behaviour and may thereby contribute to social policing in groups. Through transmitting information about others' cooperative reputations individuals can make better decisions regarding who to interact with and avoid exploitative individuals. Profitable coalitions translate to fitness benefits hence natural selection should promote cognitive mechanisms that allow for quick detection, memory storage and effective reconstruction of reputational information. Here we investigated differences in the recall rate of non-social, non-reputation social, positive reputation and negative reputation information. 96 student participants were asked to listen to eight recordings of conversations involving different types of information. After a distracter task, they were asked to recall and report as much information as they could. We manipulated the perceived cooperativeness of the environment by exposing participants to cooperative or uncooperative game partners prior to the listening task. Positive and negative reputational information was recalled with a higher accuracy than non-reputational social information but the recall rate of non-social information confounded this result. Names associated with different types of conversations were recalled most accurately for negative reputation, less accurately for positive reputation and the least accurately for non-reputation social and non-social information. Exposure to cooperative and uncooperative environments did not affect recall rates. We discuss the results with reference to the Machiavellian intelligence/social brain hypothesis and the debate over human cognitive sensitivity to cheaters versus sensitivity to the rarer cooperative type in the group.

6.1. Introduction

'No one gossips about other people's secret virtues.'

Bertrand Russell

In everyday life people often rely on third-party information relating to evolutionarily important matters such as resource acquisition, diseases or potential sexual and social partners. About two thirds of conversation time is devoted to social topics which suggests that people value information about others' personal and social lives and are interested in spreading such information about themselves (Dunbar et al., 1997). Moreover, cultural transmission of social information, generically labelled as gossip, is more effective than non-social information which indicates that broadcasting the former might have played an important role in human evolution (Mesoudi et al., 2006). Considering the amount of time spent gossiping and the effectiveness of gossip transmission, it is tempting to assume that gossip contributed to the evolution of human ultra-sociality. According to Dunbar (1993) complex social networks enforced the increase of neo-cortex size and the evolution of language which allowed for a high level of cooperation in human groups.

The role of gossip in the evolution of human cooperation among unrelated individuals rests on two functions of language: bonding and social policing. It has been proposed that language in humans bears functional similarities to grooming in non-human primates (Dunbar, 1996). By enhancing social bonding, verbal communication positively affects cooperation even if individuals are strangers and when it pays to be selfish. In experimental games, levels of cooperation increase dramatically when communication is allowed (for a review see: Ostrom, 2003, and Sally, 1995). Discussing optimal strategies could potentially account for higher cooperation in groups who can chat; but cooperation levels seem to be enhanced even more by physical closeness. Experiments show that what really matters in solving social dilemmas is the modality in which information is transferred and whether the communication occurs via computer or face-to-face (Jensen et al., 2000, Bicchieri and Lev-On, 2007).The fact that the physical contact during the transmission affects cooperation stronger than the content of the transmitted information emphasizes the bonding role of gossip in the development of pro-social behaviour.

In complex social groups consisting of over 150 individuals, cognitive skills do not allow for directly following everyone's reputation and direct social policing (Dunbar, 1993). The stability of such groups is potentially endangered by free riders who exploit others and move away to find an inexperienced victim. However, when information about free riders' reputation can be spread, their success decreases and cooperation becomes stable (Enquist and Leimar, 1993). It has been postulated that the main role of gossip in modern societies is managing reputations (Emler, 1990). Although, when gossiping, people devote only about 3% - 4% of time to spreading negative opinions about others (Dunbar et al., 1997); evidence from economic experiments suggests that the possibility of gossip successfully discourages people from free-riding and promotes cooperation (Piazza and Bering, 2008). In order to use gossip as a policing tool, third-party information needs to accurately reflect individuals' behaviour. It has been shown that information obtained from others can be a reliable substitute for direct observation (Sommerfeld et al., 2007). However, gossip is prone to manipulation and for this reason humans use a set of assessment techniques to verify its veracity (Hess and Hagen, 2006, Sommerfeld et al., 2008). In further support of the policing function of gossip it was demonstrated that people disapprove of self-serving gossip but approve of gossiping in response to norm violation (Wilson et al., 2000). If gossip plays such a significant role in social control, information about reputation should be transmitted with higher precision than other social information. Mesoudi et al. (2006) found that social information was passed on through transmission chains with a higher accuracy than non-social information supporting the social brain (Dunbar, 1998) or Machiavellian intelligence (Byrne, 1996) hypothesis. However, no evidence was found for the strong version of the social brain hypothesis which predicted a bias for gossip about norm-violating behaviour in comparison to everyday social behaviour information.

Policing through gossip can function successfully if people pass information about others' cooperative and selfish acts. However, it is also crucial that people store this information, so that when encountering an individual, they can easily retrieve this individual's reputation from memory and adjust their behaviour accordingly. In past research Tooby and Cosmides (1992) promoted the concept of a 'cheater detection' module in the human brain which would allow for quick identification of social norm violators. A number of studies supported this idea by reporting that humans display cognitive sensitivity to faces of cheaters (Yamagishi et al., 2003, Mealey, 1996, Chiappe et al., 2004, Verplaetse et al., 2007). Other researchers provided evidence for the existence of altruism detection ability independent of the cheater detection one (Brown and Moore, 2000, Brown et al., 2003) or a cognitive bias to cooperators affected by social context (Felisberti & Pavey, 2010). The idea of specialized areas in the brain responsible for either cooperator or cheater detection has been challenged by Barclay (2008) who showed that sensitivity to faces of cooperators and defectors depends on their frequency in the environment. Members of the group which was in the minority were more frequently recognized as previously seen and more accurately identified with regard to their pro-sociality than members of the larger group. Interestingly, in the condition with equal number of cooperators and defectors, cooperators were remembered marginally better than defectors. This finding pointed to a more general reputation-tracking mechanism that relies on focusing on the less frequent behaviour, which in human societies is defecting rather than cooperating.

A different approach to the recognition of cooperative intentions was adopted by Frank (1988). Frank considered populations with different proportions of cooperators and defectors and different cooperative intention recognition abilities. If cooperators and defectors look alike and the payoff of defection is higher than the payoff of cooperation, cooperators will become extinct. In contrast, if coopeerators and defectors are easily identified, cooperators will preferentially interact with each other and oust defectors. If defectors learn to mimic cooperators, again, cooperation will vanish. The most interesting of frank's examples is when there is a cost of a scrutiny of cooperative intentions. Frank noticed that engaging in such costly examination of intentions would depend on the proportion of cooperators and defectors in the population. The proportion of cooperators must be sufficiently low for the costly examination to occur, because with a high proportion of cooperators an individual will often interact with them by chance. Only when the cost of scrutiny is lower than the cost associated with frequent interactions with defectors, will such scrutiny take place. It is worth noting that according to Frank it will never pay for defectors to costly scrutinize others. This is because, with a low number of cooperators who assess cooperative intentions, a defector will be scrutinized and rejected as a partner.

A cognitive bias towards detecting cheaters or cooperators in the environment could also be modulated by individuals' own cooperative behaviour. Exploitative individuals might, for example, be more aware of the possibility of being exploited by others than pro-social individuals and for this reason might focus more on cheaters than cooperators. It has been shown that people can be divided into three stable and distinct cooperative types: spiteful, payoff-maximizing and altruistic (Kurzban and Houser, 2005, Simpson and Willer, 2008). Pro-social inclinations are measured using the Social Value Orientation scale (SVO) which categorizes individuals into competitors, pro-selves and pro-socials (Van Lange et al., 1997). Whether individual pro-social predispositions affect the cognitive decisions involved in detecting cheaters and cooperators is currently unknown.

The present study was inspired by the concept that human cognitive skills evolved to solve social problems that occur in complex groups (Mesoudi et al., 2006, Dunbar, 2004); and by the debate over the cognitive mechanisms involved in social policing (Barclay, 2008). We first intended to test the strong version of the social brain hypothesis which assumes that information about norm violators is more likely to be attended to and transferred than other types of information. Mesoudi et al.'s (2006) findings support the soft version of this hypothesis – bias in transmitting social in comparison to non-social information. However, the researchers did not find that gossip-like information (describing an affair of a student with her professor) was transmitted with a higher accuracy than non-gossip social information as predicted by the strong version. The content of Mesoudi et al.'s example might not have tapped the exploitative nature of human social relationships as it referred to sexual rather than cooperative reputation. To test the strong version of the social brain hypothesis we investigated the recall rates for conversations involving non-social, non-reputational social and social reputational information in which reputations referred to an individual's pro-sociality and fairness. We predicted that information about cooperative and uncooperative acts and names of actors performing them would be recalled with higher accuracy than all other types of information and the associated names.

Secondly, we addressed the question of whether humans are more sensitive to gossip about cheaters or cooperators or whether the sensitivity is contingent on the experience one is exposed to. If humans are specialized cheater detectors they should be able to recall the information regarding the negative reputation and the names of actors performing uncooperative acts with higher accuracy than other kinds of information. If, however, people use the more parsimonious 'focusing on the rarity' rule they should selectively attend to and memorise information about the group that is in the minority (Barclay, 2008). In this case, we would predict that people exposed to an uncooperative environment would recall information about cooperative acts and the names of actors performing them with higher accuracy than information and names associated with cheating and *vice versa.* An alternative prediction follows from Frank's (1988) frequency-dependent costly scrutiny. According to Frank with a high proportion of cooperators no sensitization to cooperative intentions should occur. When the proportion of cooperators is low, they should become able to distinguish cooperators from defectors.

To determine which mechanism people use to assess others' cooperative intentions we primed participants with a cooperative, uncooperative or neutral environment, after which we exposed them to conversations involving gossip about positive and negative reputations of others, and measured the recall rate. This study served as a critical test between the two competing hypotheses, cheater detection and rarity detection, and aimed to determine which of them was correct. We also administered the SVO scale in order to test whether individual pro-social predispositions affected recall rates (Van Lange et al., 1997), but because this analysis was exploratory we did not make any specific predictions.

6.2. Methods

6.2.1. Participants

86 female and 10 male student participants (M age = 20.22, SD = 1.87) were recruited at Newcastle University via e-mail. The study was directed to English native-speakers and advertised under the title 'Intensive listening and performance in simple tasks'. Participants completed it individually and were rewarded with module credits and small payments based on their performance in various tasks.

6.2.2. Material

After the priming stage (described in the Design and procedure section), participants listened to eight conversations which represented four conditions (two conversations per condition). The conversations contained (1) non-social information defined as a technical description not involving social interactions with others; (2) social information without reputation information (referred to as *no reputation*) which involved interactions with others but did not reveal anything about the target's social reputation (neither in terms of cooperative tendencies nor social status and sexual behaviour); (3) social positive reputation in which the described individual behaved in a pro-social or altruistic way, or (4) social negative reputation which included acts of cheating or exploiting others (see Appendix C for examples). Conversations involved different names of the main characters and were matched with regard to the number of words and propositions (units of meaning). The conversations were acted by two female English native speakers and recorded. Each conversation would start with an introductory question by Actor A e.g. 'Have you heard about Alex/Ann?' after which Actor B expressed interest e.g. 'What happened?' and Actor A replied by presenting the target story. Conversations were presented in a balanced order. We used eight different orders in which both the sequence of conversations and the sex of the characters were varied. Two conversations of the same type were never presented in a consecutive order. The main character of the conversations was either a male or a female (two all male sequences, two all female sequences and four mixed sequences in which the sex of cooperators and defectors was balanced) with the exception of the non-social conversations (involving a technical description of an activity by Actor A) in which only females were featuring.

The collected scripts with recalled conversations were compared to the original ones. We calculated the number of correctly recalled propositions in each conversation (i.e. all relevant units of meaning reported in any order). Any incorrect information reported was discarded. A proposition was defined after Kintsch (1974) as a semantic unit which conveys meaning (see also Mesoudi et al., 2006). Propositions contain predicates, which usually describe the objects and the relationships between them, and arguments, which are the objects themselves e.g. a sentence 'John gives money to a local charity' would be propositionally encoded as: gives (John, money), to a charity (money), local (charity). All target stories consisted of 10 propositions, 33-38 words and 3-4 sentences. We also collected data on the names correctly assigned to conversations. If a participant provided only a name without any reference to a conversation in which it appeared, they did not score a point.

6.2.3. Design and procedure

Participants were assigned to one of three conditions: cooperative, uncooperative or neutral environment. Those who were exposed to a cooperative or uncooperative environment played 42 rounds of a Prisoner's Dilemma game (PD) with a virtual partner who either cooperated in 80% or 20% of rounds. In PD participants could decide whether they wished to cooperate or defect. Participants received the lowest payoff if they cooperated and their partner defected, a higher payoff when both players defected, an even higher payoff when both cooperated and the highest payoff when they defected and their partner cooperated. The points earned in this task translated to real money which, together with earnings from other tasks, were summed after the experiment. Participants received an e-mail informing them about the amount of money they earned and asking to arrange an appointment with the experimenter to collect the payment.

Participants in the control condition (neutral environment) were asked to do 20 mathematical calculations which took a similar amount of time to playing PD. Next, participants were asked to listen to eight conversations (each lasted between 16s and 19s), of which some involved information about individuals' positive or negative reputations. Participants were asked to press 'Play' when they were ready to listen to a conversation and 'Continue' after they finished in order to proceed to the next conversation.

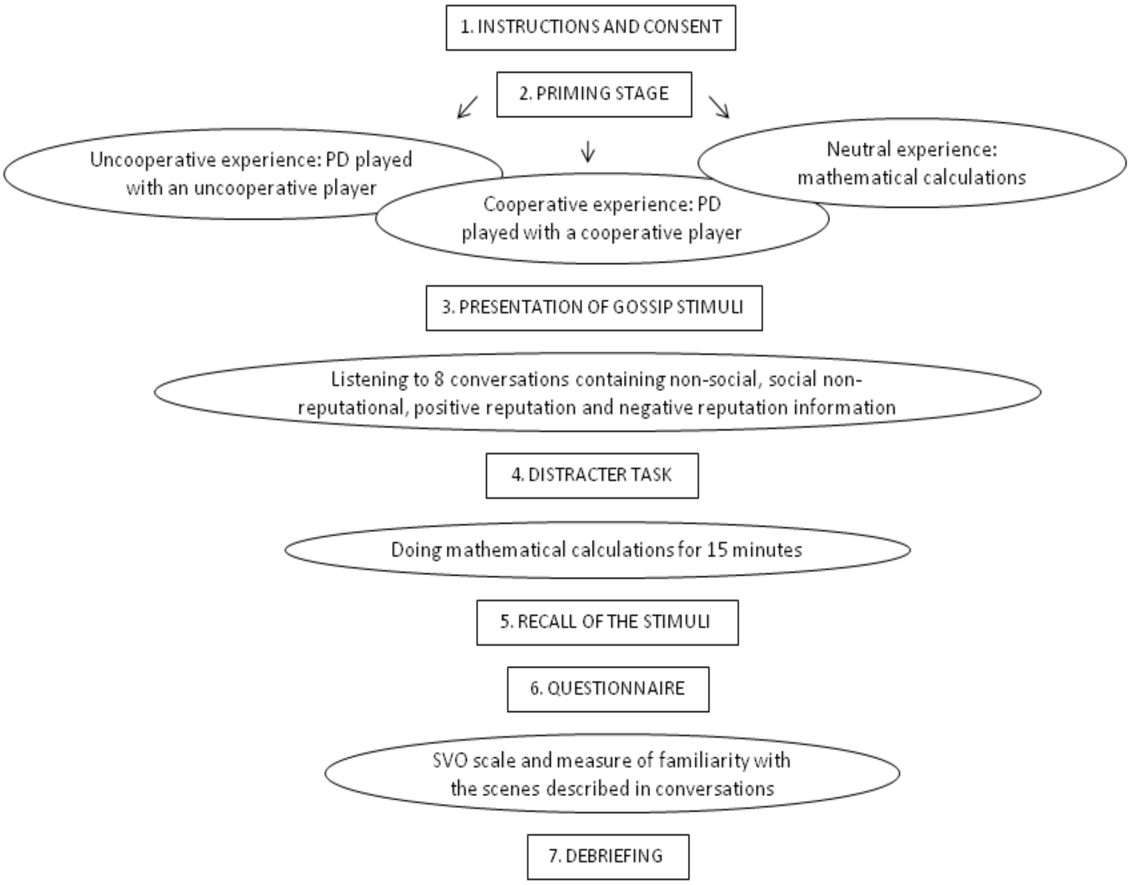

1. INSTRUCTIONS AND CONSENT
2. PRIMING STAGE
Uncooperative experience: PD played with an uncooperative player
Neutral experience: mathematical calculations
Cooperative experience: PD played with a cooperative player
3. PRESENTATION OF GOSSIP STIMULI
Listening to 8 conversations containing non-social, social non-reputational, positive reputation and negative reputation information
4. DISTRACTER TASK
Doing mathematical calculations for 15 minutes
5. RECALL OF THE STIMULI
6. QUESTIONNAIRE
SVO scale and measure of familiarity with the scenes described in conversations
7. DEBRIEFING

Figure 6.1 Diagram showing different stages of the experiment.

After listening to the conversations participants were distracted by doing complex calculations for 15 minutes. Then, they were asked to recall as much as they could from the eight conversations that they heard and write it down in any order in the eight boxes provided on the same webpage (the boxes were not labelled in any way). Participants were encouraged to take as much time as they needed to complete this task. After that, participants completed the SVO scale and a questionnaire in which they had to indicate their familiarity with the situations described in the conversations i.e. they were asked to indicate on a Likert scale how often they perform or see someone else performing the activities described in conversations (e.g. the response to a question how often a person donates blood ranged from never to regularly). The study was computerized using the Qualtrics software for questionnaires (Qualtrics Labs Inc., Provo, UT) and z-Tree for the PD game (Fischbacher, 2007). For a diagram of the procedure see Figure 6.1.

6.3. Results

Overall, the recall task proved to be relatively difficult: participants managed to recall at least one piece of information from on average 4.68 conversations (out of 8 conversations, SD = 1.55). Scores of the recall rates from two conversations of the same condition were pooled into one variable.

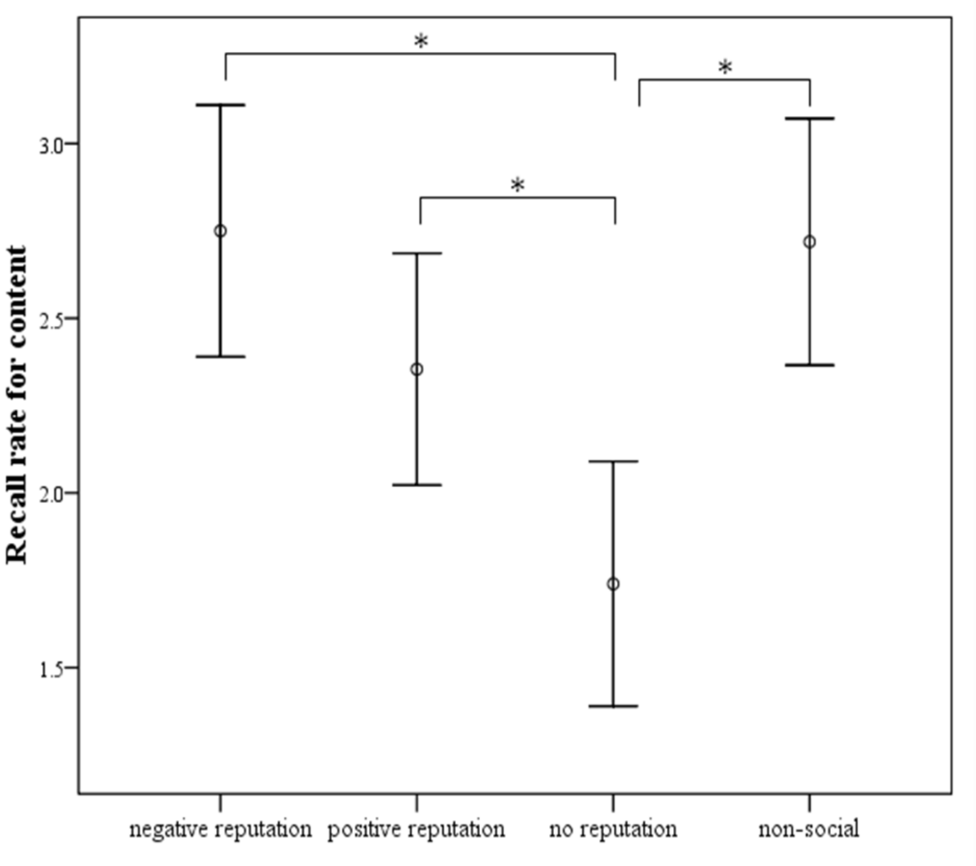


Figure 6.2 Mean recall rate for conversation content with 95% confidence intervals by condition. The maximum score for content in each scenario was 10.* p < 0.012

The distribution of the variables tested was non-normal and none of the transformations improved the normality. We decided to use a mixed ANOVA model because it is relatively robust to departures from normality in particular when the assumption of homogeneity of variances is satisfied as in our data (Howell, 2002). A mixed 4(conditions: negative reputation, positive reputation, no reputation, non-social) × 3(environment: uncooperative, cooperative, neutral) ANOVA was applied to the content recall and then to name recall. We found a significant effect of condition on the rate of content recall, F (3, 279) = 8.83, $p < 0.001$ (see Figure 6.2). To further investigate the effect of conversation content we did planned comparisons between the different conditions. We used a Bonferroni corrected significance of $\alpha^* = \alpha/N$ where N = number of comparisons (hence $\alpha^* = 0.05/4 = 0.0125$). Content recall accuracy was significantly different from other conditions only in the *no reputation* condition (Figure 6.2).

For each recalled name correctly associated with a conversation participants could score a point. We averaged scores from each two names of the same condition, so the maximum score for each condition was 1. ANOVA revealed a significant effect of condition on the rate of name recall, F (3, 279) = 17.88, p < 0.001 (Figure 6.3). Name recall accuracy was the highest in the *negative reputation* condition, lower in the *positive reputation* condition and the lowest in the *no reputation* and *non-social* conditions (Figure 6.3).

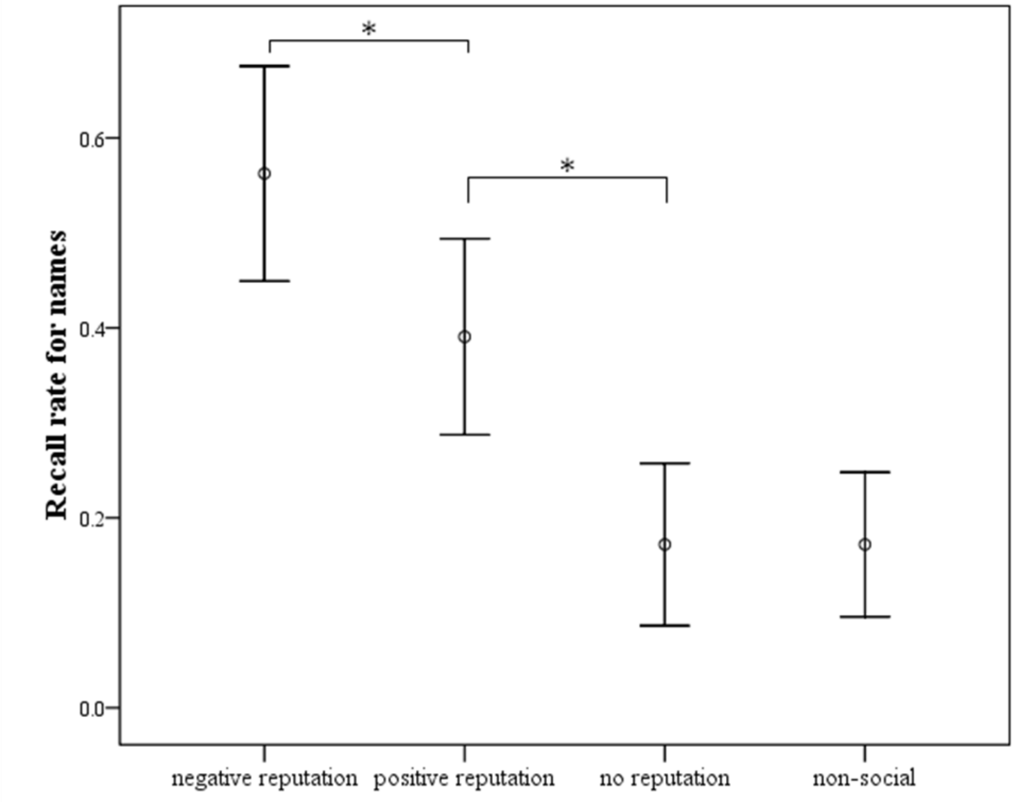


Figure 6.3 Mean recall rate for names with 95% confidence intervals by condition. Participants were assigned a score of 1 if they recalled a name and associated it with the appropriate scenario. * p < 0.012

Environment affected neither the content recall rates, F(2,93) = 2.19, p = 0.18, nor the name recall rates, F(2,93) = 0.29, p = 0.75. There were also no interactions between condition and environment either for content recall, F(6,279) = 1.40, p = 0.21, or for name recall, F(6, 279) = 1.34, p = 0.24. We found no relationships between the recall rate of conversations and familiarity with the discussed topic (all correlations non-significant).

In the second step of the analysis, based on SVO scores participants were assigned to pro-social, pro-self and competitive types if they made six or more consistent choices of money distributions (for details see: Van Lange et al., 1997). A mixed 3(SVO type)×4(condition) ANOVA was conducted using data from 20 competitive, 24 pro-self and 17 pro-social individuals (the data of participants who provided inconsistent choices were removed). There was no main effect of the SVO type on the content recall rate, F (2, 58) = 0.62, $p > 0.05$ and no interaction between the SVO type and recall condition, F (6, 174) = 0.81, $p > 0.05$. With regard to name recall rate, we found a main effect of SVO type F (2, 58) = 3.91, $p < 0.05$ (see Table 6.1 for descriptive statistics), but no interaction between SVO type and condition, F (6, 174) = 1.19, $p > 0.05$. Competitive types (M = 0.46, SD = 0.34) recalled significantly more names in general than pro-social types (M = 0.20, SD = 0.22), t(35) = -2.59, $p < 0.05$.

Table 6.1 Means and standard deviations of name recall rate by SVO type.

	SVO type		
	Pro-self	Pro-social	Competitive
Negative reputation	0.69(0.55)	0.44(0.50)	0.65(0.61)
Positive reputation	0.35(0.45)	0.18(0.35)	0.62(0.56)
No reputation	0.25(0.62)	0.03(0.12)	0.30(0.44)
Non-social	0.06(0.22)	0.18(0.35)	0.25(0.50)

6.4. Discussion

The fact that humans spend a considerable amount of time discussing social topics and the evidence that social information is transmitted with higher efficiency than non-social information support the social brain hypothesis (Dunbar et al., 1997, Mesoudi et al., 2006). However, if language, and in particular, gossip, serves to enhance social policing in complex groups, information about reputations should be attended to more than other types of social information. The social brain or the Machiavellian Intelligence hypothesis emphasizes the manipulative potential of human cognition; hence, it is reasonable to expect that people evolved certain adaptations to allow them to indirectly monitor others (Dunbar, 1998, Whiten, 2000, Byrne, 1996). Here we show that human memory is indeed sensitive to gossip about others' reputations. We found support for the stronger version of the social brain hypothesis, namely, conversations containing reputational information and the names of featuring characters, in particular the ones who violated social norms, were recalled with a higher accuracy than conversations with non-reputational social information. This result is in line with the findings of Wyer Jr et al. (1994) who observed that participants recalled unfavourable behaviours of others which were mentioned in conversations better than favourable behaviours.

Participants in our sample were exceptionally accurate at recalling the names of cheaters. In order to make a profitable decision of whether to engage in a relationship with someone, an individual does not need the entire detailed history of their cooperative behaviour; a parsimonious way of assigning reputation is to label an individual according to how they behaved in a recent interaction e.g. 'good' or 'bad' and use a simple rule: avoid 'bad' (see the literature on image scoring e.g. Nowak and Sigmund, 1998). Remembering names together with their reputation appears to be more relevant from an evolutionary perspective than storing information about the details of others' good or bad deeds. People's faces are the most distinctive individual characteristics hence it is reasonable to assign reputations to them (Verplaetse et al., 2007). However, when no faces are available e.g. when gossiping about unknown and unseen persons or when interacting online, names allow for quick identification of individuals. Our results indicate a cognitive sensitivity to names associated with reputations; thanks to this mechanism norm violators can be easily identified and avoided even without prior direct interaction with them.

Our results do not support Mesoudi et al.'s (2006) findings in that the recall rates for social non-reputational information are not higher than for non-social information; in fact they are significantly lower. The lack of consistency between the two studies could be explained by the differences in the material used. In our study all material was presented in a conversational form even if it involved a non-social, technical description of some activity, while Mesoudi et al. presented the material as stories which were read by participants. The mere fact that two individuals exchanged information could have in itself made the technical content be perceived as social. Further, providing another individual with desired non-social information such as giving a recipe for a tasty cake or instructions how to fix a puncture (examples used in the conversations) might have been perceived as pro-social behaviour and might have increased the speaker's reputation and hence be recalled with a high accuracy. Alternatively, the conversation content in the non-social information condition could have been easier to reconstruct than other types of information that is: if someone remembered that the conversation concerned fixing a puncture, one could easily add what one knows from experience about this activity and by doing so report more information than was actually stored in memory from the conversation. The recall rate for names, which was lower for non-social than for reputational information, supports such an explanation.

High recall rate and accuracy for names associated with norm violators suggests that human cognitive skills are socially fine-tuned in order to minimize the risk of exploitation. The threat of spreading negative gossip can successfully facilitate cooperation especially when the potential gossiper can identify the subject of gossip (Piazza and Bering, 2008). Our result does not mean however that gossip about negative reputations should be trusted more because of its veracity. As skilled Machiavellians people can manipulate gossip to defame others. Sommerfeld et al. (2008) showed that when presented with a few negative gossips about a potential partner, the variance in cooperative behaviour towards this partner was higher than when the gossip involved mixed positive and negative information. This indicates that despite the powerful impact of negative gossip on reputation, people do take into account the possibility of others acting spitefully and transmitting false information.

Another objective of this study was to test between two competing hypotheses regarding the mechanism people use to effectively follow others' reputations by means of gossip. We did not find support for the idea that the cognitive sensitivity to reputational gossip depends on the cooperativeness of one's environment, as was shown by Barclay (2008) for face stimuli and as was postulated by Frank (1988). Priming with playing against a cooperative or uncooperative partner did not have any effect on the gossip recall rate. Our results are therefore more in line with the notion of the cheater detection module according to which sensitivity to norm violations is hard-wired and occurs regardless of individuals' cooperative experience (Cosmides and Tooby, 1992). Although a general rule of attending to the rarer group would be a more cost-effective mechanism (Barclay, 2008), it is not likely that humans in their evolutionary history have ever lived in groups where the majority of group members were cheaters. Extensive cooperation among group members occurs even in 'simple' hunter-gatherer societies which suggests it might have appeared very early in human evolutionary history (Hill et al., 2009). Moreover, it is probable that even if at some point cheaters started to dominate in a group, cooperative individuals would have decided to 'walk away' and move to a more cooperative group (Aktipis, 2004). It has been shown that people tend to voluntarily move from a non-sanctioning and not very cooperative group to a group in which free-riders are punished and as a consequence cooperation is higher (Gürerk et al., 2006).

Our results indicate that an individual's pro-social orientation is not related to recall of specific reputational information. However, we found that, overall, competitive types performed better at recalling names than pro-social types. Competitive types are not just selfish, but strive to maximize the positive difference between their own and others' gain. Therefore, they might be more sensitive to the identities of individuals they compare themselves to. Unfortunately the present data do not allow for a more in-depth interpretation of this result and further research on is this topic is needed.

The plethora of studies on indirect reciprocity (e.g. Leimar and Hammerstein, 2001, Milinski et al., 2002b) emphasizes the importance of reputations in human cooperative behaviour. The majority of these studies, however, focus on actions towards individuals observed to behave in a certain way. With large group sizes it is not possible to follow others' reputations by means of observation. Language enabled a more sophisticated and robust form of monitoring through passing reputational information. Gossip, therefore, may have played a crucial role in the evolution of human cooperation and deserves scientific investigation. Our study shows that reputational information is recalled with a greater accuracy than social non-reputational information and that the names of norm violators are attended to more than names associated with other contexts. Our results support the strong version of the social brain/Machiavellian Intelligence hypothesis by showing that people have a memory bias for reputational gossip, in particular, gossip about norm violations and that this bias is not affected by the preceding experience of being exploited or helped.

7. The role of Theory of Mind in assessing trustworthiness

People vary in the extent to which they can assess others' trustworthiness. We investigated whether Theory of Mind (ToM), the ability to represent mental states of others, is related to accuracy in recognizing cooperative intentions. Participants completed tasks measuring social-perceptual and social-cognitive ToM and were asked to assess photographs of people playing Prisoner's Dilemma games taken at the very moment when they were making a decision to cooperate or defect. We found no relationship between ToM and cooperative intention recognition. Surprisingly, in contrast to previous studies, participants in our sample performed poorly in the trustworthiness assessment tasks. Our results question human expertise at identifying cheaters and cooperators and indicate a lack of association between ToM and trustworthiness assessment. The findings are discussed from the perspective of an evolutionary arms race between reading and masking cooperative intentions.

7.1. Introduction

Ultra-sociality and large scale cooperation towards unrelated individuals have been identified as potential driving forces behind the evolution of human-specific cognitive machinery (Dunbar, 2003, Moll and Tomasello, 2007, Hill et al., 2009). According to Dunbar (2003), increased group size and, in consequence, more complex social interactions often involving encounters with strangers put pressure on human cognitive capacities. Not being able to directly observe other individuals' actions creates a problem of how to keep track of free-riders. Free-riders undermine the stability of social systems by reaping the benefits without incurring any costs in cooperative interactions. The need to detect free-riders and maintain high levels of cooperation could explain the existence of language (Dunbar, 1996), some of the pro-social emotions (Price et al., 2002), and socially oriented reasoning (Cosmides and Tooby, 1992).

Gathering reputational information about a potential partner can aid in predicting their cooperative intentions, assessing their trustworthiness and making a decision whether to engage in an interaction or not. However, individuals often meet strangers whose reputations are not known. In such circumstances, the only way to assess someone's trustworthiness is to read subtle cues of cooperative intentions from a face or interpret non-verbal body language. Evolutionary research shows that faces reveal important information about potential mates and social partners (e.g. Rhodes, 2006, Todorov et al., 2008). Decisions about who to trust are affected by stable facial features e.g. attractiveness, similarity to kin or facial width (for a summary see Stirrat and Perrett, 2010). People also use others' facial expressions to determine cooperative intentions and, as reported by Verplaetse and colleagues (2007), after viewing photographs of individuals who played a Prisoner's Dilemma (PD) game, can correctly guess cooperative intentions with a probability higher than chance. In PD games players make simultaneous decisions whether to cooperate or defect. In a situation when one defects and the other one cooperates, the defector gains the maximum payoff while the cooperator receives a very small payoff or nothing. Therefore, individuals are tempted to defect by the high potential payoff and by the risk of being exploited.

Theoretically, there could be two opposing evolutionary pressures acting on human cognition: one promoting cheater recognition and another one favouring masking uncooperative intentions (see Hanley et al., 2003). In fact, signals of cooperation might evolve to be deceptive in a similar way as it occurs in the mating context in animals e.g. some male crickets instead of a nutritionally valuable nuptial gift may offer a female an empty silken balloon (Maynard Smith and Harper, 2003). The co-evolution of trustworthiness detection and disguising uncooperative intentions would result in an overall low ability to predict cooperative intentions (Dawkins and Krebs, 1979). However, a number of recent studies suggest that, despite a variation in the ability to predict others' cooperative behaviour, people perform at least better than chance when assessing others' trustworthiness (e.g. Verplaetse et al., 2007, Oda et al., 2009b, Fetchenhauer et al., 2010, Oda et al., 2009a, Brown et al., 2003, Frank et al., 1993) Could this variation be explained by between-individual differences in the Theory of Mind (ToM)?

ToM is one of the dimensions of social intelligence and refers to the ability to read others' minds i.e. understanding and interpreting mental states of others. It consists of at least two components subserved by different neural mechanisms (Sabbagh, 2004). The social-perceptual component involves reading facial or body cues and from them representing others' thoughts and desires. In the classic task testing this skill participants have to visually assess an individual's mental state from a photograph of their eye region (Baron-Cohen et al., 2001). The social-cognitive component, on the other hand, describes the capacity to infer about the reasoning of others e.g. "I suppose he thinks…". Social-cognitive ToM can be represented hierarchically by using different levels of social embeddedness e.g. "I understand that you want me to believe that he thinks…". The standard task measuring the social-cognitive component involves reading or listening to stories about characters socially interacting with each other. Participants are then asked to answer questions about the characters' beliefs at different levels of social embeddedness (Stiller and Dunbar, 2007).

Are there any grounds for expecting a positive relationship between ToM and the ability to assess cooperative intentions? A person with high ToM skills, by definition, should be able to infer about others' mental states pertaining to cooperative behaviour. The social-perceptual component of ToM appears to capture recognition of facial cues of trustworthiness particularly well. Cooperation and defection invoke certain emotions such as gratitude, liking, nervousness, shame or anger. Hence, the ability to recognize such emotions correctly might help in determining someone's cooperative intentions. Concealed emotions can be manifested as microexpressions lasting for 1/25-1/5 of a second (Ekman and Friesen, 1969) or slightly longer inconsistent emotional expressions (Porter and Brinke, 2008). The proficiency in recognising emotions in general may translate to spotting any false or inconsistent emotions and, in consequence, the willingness to cheat. Alternatively, another cue of trustworthiness could be emotional expressiveness itself: cooperative individuals display more both positive and negative emotions (Boone and Buck, 2003, Schug et al., 2010). Predicting other's cooperative behaviour could also be related to the social-cognitive component of ToM. In this case, however, it is more likely that individuals of high ToM skills would make more accurate trustworthiness judgements based on third-party information (gossip) rather than on facial cues.

No previous studies have examined a possible relationship between ToM and trustworthiness recognition but some links have been found between ToM, pro-sociality and Machiavellianism. In preschoolers, children with more developed ToM (measured by the false-belief Sally-Anne task) were shown to have higher preferences for fairness (Takagishi et al., 2010). In adults, social-cognitive but not social-perceptual ToM positively correlated with the personality dimension of Agreeableness (Nettle and Liddle, 2008). Agreeableness reflects inter-individual differences in concern for others and highly agreeable people are considered as friendly, warm, cooperative and helpful (Graziano and Eisenberg, 1997). Paal and Bereczkei (2007) found a moderate positive relationship between the social-cognitive ToM and pro-sociality, but no link between ToM and Machiavellianism (manipulative and exploitative strategy). Ali and Chamorro-Premuzic (2010) observed a negative association between social-perceptual ToM and Machiavellanism. In a similar vein, a negative relationship between both ToM components and Machiavellianism was reported by Lyons, Caldwell and Schultz (2010) who concluded that the need to manipulate and deceive conspecifics was not one of the driving forces behind the evolution of human social intelligence. It is possible that mind-reading ability in humans evolved not because it aided in deceiving others, but because it was needed in assessing trustworthiness of potential cooperation partners.

Another interesting question relating to the assessment of cooperative intentions is whether individuals' own pro-sociality is associated with the accuracy in predicting cooperation in others. Oda et al. (2009b) suggested that, because altruists risk being exploited, their ability to accurately predict cooperative intentions of others should be higher than the ability of defectors. A similar prediction was made by Naganawa et al. (2010) who incorrectly linked the concept of 'greenbeard effect' to human cooperation in expecting that "*altruists can detect altruists easier than non-altruists*" (p.2). The greenbeard effect assumes preferential assortment but not preferential recognition of individuals carrying the same cooperative gene (West et al., in press); the assortment can be facilitated e.g. by individuals with the cooperative gene occupying the same environment (Hamilton, 1975). Detecting altruistic individuals can be as beneficial to altruists as to cheaters, because both groups can maximize their payoff by interacting with altruists. Therefore, even if altruists could convey some cues of their behaviour via faces, there would be as strong an evolutionary pressure on cheaters as on other altruists to learn to detect them. In support of such reasoning, neither Oda et al. (2009b) nor Naganawa (2010) found any association between individuals' own altruism and the accuracy of recognizing altruism in others.

This study examined the possible role of social intelligence in trustworthiness assessment. We administered both the social-perceptual and social-cognitive ToM measures and additionally examined participants' pro-social orientation. Participants were presented with photographs of people who cooperated or defected in a PD game and were asked to guess their decisions. We predicted that social-perceptual ToM would be positively associated with cooperative intentions recognition. Based on research highlighting human sophistication in guessing cooperative intentions, we expected that participants would be able to correctly assign cooperative intentions with a probability higher than chance. Finally, considering Oda's et al. results (2009b) we predicted that more cooperative individuals would not be more accurate at identifying cooperative intentions.

7.2. Method

7.2.1. Participants

We collected data from 100 students: 15 males (mean age = 21.6, SD = 3.90) and 84 females (mean age = 19.7, SD = 2.41); the sex and age of one participant was unknown. In the analysis only data from English native speakers or non-native speakers who spent at least one year in the UK were used (99 participants). Students were asked to do the study in a computer cluster after they finished their class in research methods. The tasks, presented in a random order using Qualtrics survey software, took approximately 40 minutes to complete and the students were rewarded for their time with course credits.

7.2.2. Materials

Social-cognitive ToM task - an updated version of the task used by Stiller and Dunbar (2007)

Participants were asked to listen to a set of five short stories describing social situations (e.g. about a woman trying to receive a wage increase from her boss) and answer ten memory questions (true or false) after each story was presented. Five questions referring to different levels of embeddedness were mixed with five questions about the factual content of the story. ToM questions involved between two (e.g. 'Emma wanted more money.') and six (e.g. 'Emma believed that Jenny hoped that her boss, the greengrocer, would believe Emma's claim about the chemist wanting to offer her a job.') levels of embedding. The questions pertaining to factual events also involved different levels of complexity and were included in order to control for the participant's understanding of complex sentences. The stories were recorded by a professional actor in a sound-proof cabin. For analysis we calculated the number of correctly answered ToM and factual questions in all stories.

Reading the mind in the eyes test (Baron-Cohen et al, 2001)

Participants were required to match each of 36 pictures of pairs of eyes to one of four words depicting complex emotions. Participants were provided with instructions, including a glossary for the terms used to describe the emotions. Each correct response scored a point. The test has been used as a measurement of affective ToM capacity in both clinical and non-clinical populations (Baron-Cohen et al., 2001).

Assessing trustworthiness of faces (Verplaetse et al., 2007)

Participants were presented with a set of 26 photographs used as stimuli in a previous study (Verplaetse et al., 2007, see Appendix D for examples). The photographs depicted faces and torsos of Belgian students who had played a PD game. The photographs were taken at the very moment when the students were making a decision to cooperate or defect. Participants were first familiarized with the PD game and had to pass a comprehension test in order to proceed. Then, the photographs of 13 cooperative (nine male and 4 female) and 13 uncooperative (seven male and six female) faces were presented to them in a random order, each accompanied with a question asking whether the photographed person cooperated or defected.

Social Value Orientation Scale (SVO) (Van Lange et al., 1997)

Participants were asked to choose one of the three presented options of sharing a sum of money between them and another person. The options included a fair share, and two unequal shares, one in which a participant would receive the highest payoff in comparison to other options, and another one in which the positive difference between a participant's and the other person's payoff would be highest. The options represented respectively the pro-social, pro-selfish and competitive orientation. Participants had to make a decision nine times, each time being presented with different values in a different order. Participants were classified as: cooperators, individualists or competitors depending on the number of choices made in each category. Only if participants' choices were consistent (at least seven choices of the same category) were they included in the analysis of the effect of SVO on ToM skills or trustworthiness assessment. For other analyses, sample sizes of other variables were not reduced to the value of sample size for SVO.

7.3. Results

In order to satisfy the assumption of normal distribution scores for trustworthiness assessment were square rooted and scores for the Eyes test were cubed. Scores for the social-cognitive and social-perceptual component of ToM were significantly higher than chance, t(98) = 29.51, $p < 0.001$ and t(98) = 20.11, $p < 0.001$ respectively (see Table 7.1 for descriptive statistics). Scores for trustworthiness assessment differed from chance but in an unexpected direction: participants accurately identified cooperative intentions with a probability significantly lower than chance, t(98) = -2.01, $p < 0.05$. Theoretically, it is possible that people have a bias in categorising others as trustworthy or untrustworthy irrespective of the actual cooperative status of the person who is being judged. In order to test whether such a bias existed in our sample, a one-sample t-test was conducted to determine whether individuals were more likely to categorise others as cheaters than cooperators. Participants, who had not known that there was the same number of cooperative and defecting faces in the presented set, were not biased to categorise faces as cheaters or cooperators, t(98) = 0.61, $p > 0.05$.

Anecdotally, women are more pro-social, charitable and empathic than men but a recent study showed that male and female pro-social behaviour depends on the stakes involved and that the question of which sex is fairer cannot be easily answered (Andreoni and Vesterlund, 2001). Nevertheless, we tested whether male and female faces are perceived differently. Due to unequal sex distribution in the stimuli we calculated the proportion of male and female faces identified correctly i.e. the correct score for each sex was divided by the total number of photographs of each sex. We then compared these proportions with a Wilcoxon test and found no difference in accuracy when judging male (Med = 0.5) or female (Med = 0.5) faces, z(98) = -0.44, $p > 0.05$. We also investigated whether there is any bias to perceive women as more pro-social. Again, we calculated the proportions of female and male faces identified as cooperative (irrespective of whether the identification was correct) and compared these proportions with a Wilcoxon test. We found that female faces (Med = 0.6) were perceived as cooperative more often than male faces (Med = 0.44), z(98) = -7.25, $p < 0.01$ (see Figure 7.1). This result is conservative as there was a smaller proportion of female faces (4/10) than male faces (9/16).

Neither the social-cognitive nor the social-perceptual ToM score correlated with the ability to assess trustworthiness (r(97) = 0.04, p > 0.05 and r(97)= -0.04, p > 0.05 respectively, see Figure 7.2). Analogous analyses investigating the relationship between ToM scores and the assessment of (separately) cooperative and defecting faces provided the same results. Participants of different SVO categories (21 competitors, 33 individualists and 31 cooperators) did not differ with regard to how accurate they were in identifying cooperative intentions, F(2, 82) = 0.08, p > 0.05.

Table 7.1 Mean scores (M) with standard deviations (SD) for the tasks used in the study.

Task	M(SD)	Range	Maximum possible score
Social-cognitive ToM (intentionality score)	19.68(2.42)	11-24	25
Social-cognitive ToM (memory score)	20.37(1.95)	15-24	25
Social-perceptual ToM (Eyes test)	27.57(3.95)	10-35	36
Trustworthiness assessment	12.99(2.80)	8-19	26

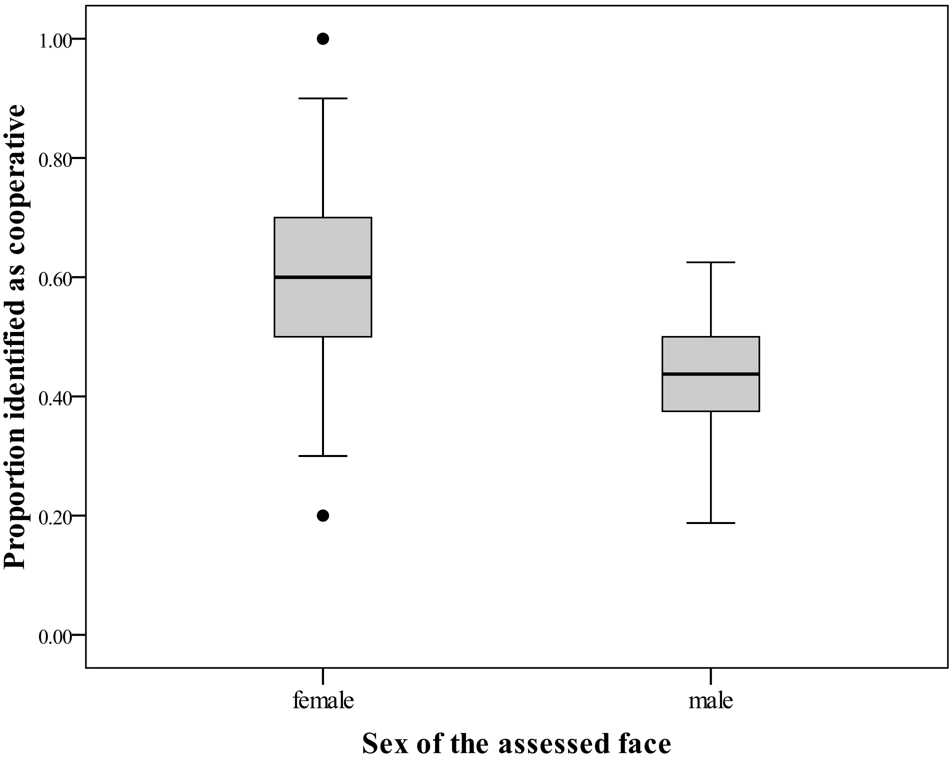


Figure 7.1 Boxplots presenting median proportions of male and female faces perceived as cooperative (with quartiles, extreme values and outliers).

We investigated whether being distrustful in general (measured by the number of faces out of 26 identified as defectors irrespective of whether the identification was correct or not) is related to ToM abilities or SVO. There was no correlation between being distrustful and either the social-cognitive ToM score (r(97)= 0.10, p > 0.05), or the social-perceptual ToM score, (r(97) = 0.02, p > 0.05). Various SVO types did not differ from each other in terms of distrustfulness, F(2, 82) = 0.17, p > 0.05.

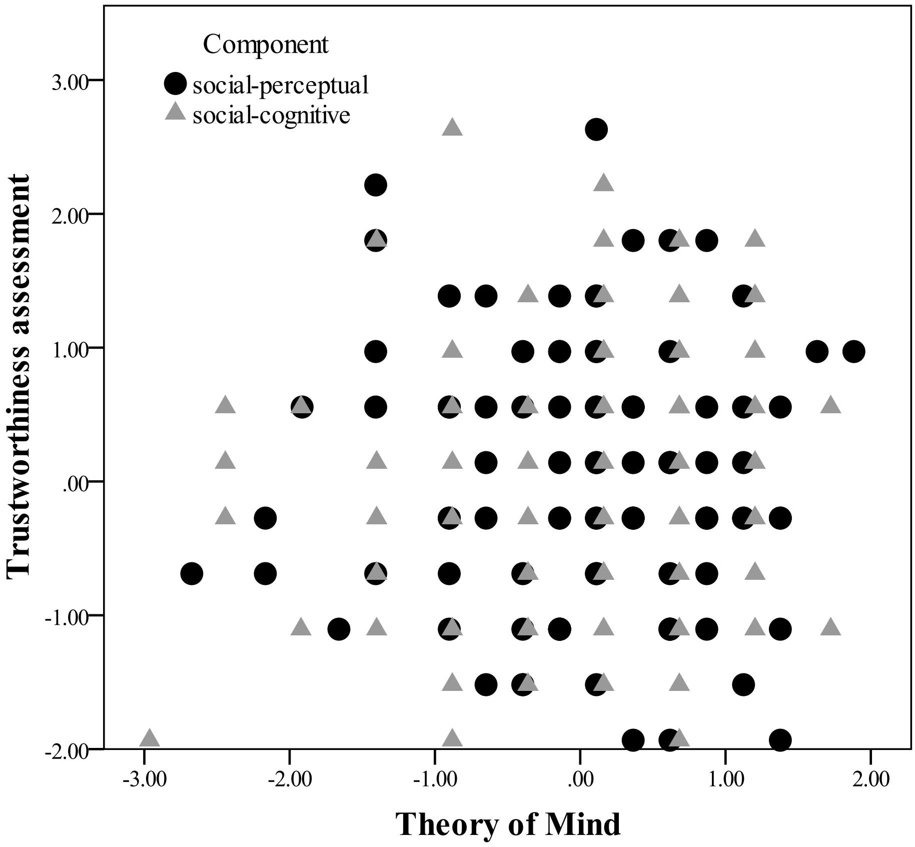


Figure 7.2 Relationship between participants' scores on ToM scales and trustworthiness assessment. For a better visual representation the values were standardized.

We can confidently reject the experimental hypothesis of the relationship between ToM and trustworthiness assessment. The likelihood of making a Type II error was small. Using a power calculator we estimated that with our sample size and the anticipated medium effect size (0.15) the probability of rejecting a false null hypothesis was over 90%.

7.4. Discussion

A number of recent reports stress human ability to predict others' cooperative behaviour from immediately available facial and bodily cues (e.g. Verplaetse et al., 2007, Fetchenhauer et al., 2010, Oda et al., 2009b). Here, we demonstrate that assessing trustworthiness is not always such an easy task. Participants did not show proficiency in identifying cooperators and defectors. Also in real life people seem to make many mistakes in predicting someone's cooperative behaviour. In a British TV show 'Golden Balls' contestants play a variation of a Prisoner's Dillemma game competing for sometimes large sums of money (up to £100 000). Data from 281 episodes of the 'Golden Balls' game demonstrate that in 44.1% of episodes one player decided to cooperate while the other one defected (Van den Assem et al., 2010). Clearly, cooperators in those pairs failed to guess what their partner would do and misinterpreted cues of defection.

Although we used the same stimuli as Verplaetse et al. (2007), participants in our sample did not identify cooperative intentions with a probability higher than chance. One possible reason for this difference is the fact that we used only event-related photographs whereas in Verplaetse et al.'s study participants could see each face in three contexts: neutral, practice round and proper round. Perhaps it is necessary for people to see how an event-related face varies from the neutral face in order to pick cues of cooperation or deception. Because humans evolved surrounded by moving and not static faces of others, being able to assess biological motion may contribute to the accuracy of distinguishing cooperators from defectors. Movement has been shown to play a role in assessing traits important in mate choice (e.g. Brown et al., 2005), so it could also affect the way in which people perceive partners for cooperative interactions. As demonstrated by Brown et al., (2003) smiling and expressions under involuntary control, which could be observed in videos, were more typical of altruists than non-altruists.

The difficulty in assessing cooperative intentions might also have been caused by the ambiguous motives for defection in the one-shot PD. An individual can decide to defect in order to gain the whole reward (in which case he deserves to be called a cheater) or because of caution and in order not to receive the sucker's payoff. Our results support other reports in which assessments of honesty in faces were not related to real honesty (Zebrowitz et al., 1996). Conceptually, under the assumption that people can accurately distinguish cooperative types, defectors should almost disappear from the population because no one would be willing to interact with them.

The main aim of the study was to examine a potential link between ToM and the ability to assess trustworthiness. More specifically, we predicted a positive relationship between the social-perceptual component of ToM and the number of correct guesses of cooperative intentions. Our data suggest that no such link exists. There can be a simple reason for this result. If, as we showed, people are not capable of picking cues of deception and cooperation, that is, if they simply guess with a probability of a correct guess lower than chance, it is not surprising that such guessing does not correlate with ToM skills.

Alternatively, the hypothesis that mind reading ability in humans evolved in order to predict cooperative intentions of anonymous individuals might still be valid under the assumption of a continuous arms race between the ability of cheaters to remain invisible and the ability of others to detect them. This arms race can be driving evolution even if at any point in time no one side is on top. High social intelligence could promote both reading and masking cues of deception. Such an interpretation would support the lack of the ability to assess trustworthiness that we report. Finally, ToM may not play such an important role in cooperative interactions as we expected. It has been shown that economic behaviour of autistic children in whom ToM skills are impaired does not differ dramatically from the behaviour of normally developing children (Sally and Hill, 2006). Perhaps the optimal strategies for playing PD and bargaining games are relatively independent of ToM skills.

As expected, and congruent with Oda et al.'s (2009b) and Naganawa et al.'s (2010) results, we found no evidence for a relationship between one's own pro-sociality and the ability to recognize cooperative individuals. The presence of such a relationship would be surprising because it would imply that the incentive for cooperators to detect other cooperators is greater than the incentive for defectors to detect cooperators. Interestingly, we found that female faces were categorized as cooperative more often than male faces. The opinion that women act more pro-socially than men was expressed early by Darwin (1871) and later supported by numerous studies (e.g. Eckel and Grossman, 1998). Although it was recently shown that the relationship between sex and cooperation is not straightforward (Andreoni and Vesterlund, 2001), perceptions of a person's cooperative intentions can be influenced by the social expectations relating to gender roles and the mass media enforcing gender stereotypes (Hyde, 2005). Our result supports that of Fetchenhauer et al. (2010) but should be treated with caution because our stimulus set contained more male than female faces.

In summary, our study questions human proficiency in recognizing cooperative intentions. The consequence of this inability may be the lack of a relationship between ToM and trustworthiness assessment. We believe our results are important in that they encourage caution when categorically asserting that humans can identify cooperative intentions. Such findings are exciting and attractive, therefore might receive more attention and publicity than non-significant results which are likely to suffer from the file drawer problem (Møllerand and Jennions, 2001, Rosenthal, 1979). Our findings should be treated seriously, considering that we had the statistical power to detect any effect. An interesting direction for future research would be to explore the arms race hypothesis which points to the co-evolution of interpreting and masking cooperative intentions.

8. Concluding remarks

8.1. Summary of results and implications

The main aim of the PhD project described in this thesis was to empirically test and explore the theory of competitive altruism (CA). Research on reputation-based cooperation has been dominated by the theory of indirect reciprocity (IR) which, as I explained in detail in Chapter 2, (1) does not account for unconditional cooperation, (2) is less likely to evolve due to its complexity and (3) suffers from problems in assigning reputational scores. In this thesis, I demonstrated that CA is a strong alternative to IR in terms of explaining cooperative behaviour. More broadly, I highlighted that reputation-based theories of cooperation could be applied to numerous instances of human interactions and therefore deserve as much attention as the traditional theories of kin selection (Hamilton, 1964) and reciprocal altruism (Trivers, 1971).

In summary, in support of CA I showed in Chapter 3 that cooperative behaviour is affected by reputational incentives. The more opportunities one has to use reputation in a profitable way, the more one is willing to invest in it. In the light of this finding, cooperation constitutes absolutely rational and strategic behaviour. In line with the assumptions of CA I also found that the initial cost of cooperation is recouped in the long-term when individuals reap benefits from the privileged access to desirable social partners and profitable interactions with them. My results concur with those of mainstream economics showing that people tend to maximize their average payoffs (Camerer, 2003). Typically, maximizing income would be a result of selfishness; however, when reputation is at stake, it actually pays to cooperate. People appear to renounce immediate benefits and maximize their payoffs by long-term rewards coming from cooperation.

In Chapter 4 I demonstrated how differences in the relative cost of reputational investments arise in the context of unequal resources. I interpreted the high relative cost of cooperation incurred by low-resources individuals as a result of the CA setting. In their paper Barclay and Willer (2007) wrote "*Competitive altruism occurs when people go beyond attempting to merely appear generous and instead actively try to be more altruistic than one another, and this has yet to be unambiguously demonstrated*" (p.749) and later *"this study provides the only unambiguous evidence to date for the existence of competitive altruism in humans and shows that partner choice is one way to produce competitive altruism"* (p.752). I would argue that the experiment presented in Chapter 4 supports the first statement to a greater extent than Barclay and Willer's study in which they only varied the reputational incentives (see Chapter 3 for details). In my experiment, the only way for low-resource participants to compete in a market for partners was to spend a substantial proportion of their allocation. In an attempt to '*try to be more altruistic than one another*', knowing about their resource handicap, the poorest participants made the costliest investments in reputation, in relative terms. The results of this study emphasize the strategic nature of human cooperation. Moreover, they unravel the mechanisms of partner choice in showing that, at least in some circumstances, people devote more attention to the relative rather than absolute cost of a cooperative investment.

In Chapter 5 I compared the efficiency of CA and IR in maintaining cooperation in social dilemmas. Previously, it has been demonstrated that the possibility of punishment enhances cooperation (Fehr and Gächter, 2002). However, results of a more recent cross-cultural study suggest that in regions with high levels of anti-social punishment maintaining cooperation via punishment is thwarted by those who punish pro-social individuals (Herrmann et al., 2008). This raises a question of how large-scale cooperation in such groups can be sustained. Reputation building is one possible mechanism to solve this problem. Milinski et al. (2002b) proposed that IR can re-establish cooperation after a decline. My results indicate that CA is a more powerful reputational mechanism in maintaining cooperation than IR. Moreover, I found that the opportunity of forming long-term partnerships positively affects cooperation.

Research presented in Chapters 3-5 stands in contrast to the view of some economists arguing that human cooperation is driven by other-regarding preferences. These economists frequently use terms such as 'altruism' and 'strong reciprocity' e.g. *"human altruism extends far beyond reciprocal altruism and reputation-based cooperation, taking the form of strong reciprocity"* (Fehr and Fischbacher, 2003, p.785). I do not deny that some form of altruism may exist. In repeated PGG games 90% of participants free-ride at the end (Camerer, 2003). Hence, 10% of participants cooperate despite others exploiting them. In Kurzban and Houser's (2005) sample 13% of participants were categorized as strict cooperators in contrast to defectors and reciprocators. The aim of the research I conducted was to explore the role of reputation-based cooperation. Although some people do cooperate no matter what, the majority is influenced by reputational incentives. Given the high levels of cooperation reputation building can evoke, it should be considered as an equally robust mechanism as the traditional theories explaining the evolution of cooperation in humans.

Considering the role of reputations in human cooperation highlighted in Chapters 2-5 it is reasonable to expect that, in comparison to non-reputational information, people (1) devote more attention to reputational information, (2) store it more effectively or/and (3) retrieve it more easily from memory. In Chapter 6 I tested whether people have any bias in the recall of reputational information. Furthermore, I investigated which of the two possible mechanisms, cheater detection or rarity detection, is more likely to shape the way people remember gossip about individuals of different reputations. My findings give partial support to the hypothesis of a cognitive bias towards norm-violators. The results were confounded by an unexpected finding suggesting that people can recall as much information about non-social conversations as about conversations involving negative reputations. However, the recall rate of names associated with different actions fitted the trend expected under the assumption of the cheater detection hypothesis. My results suggest that people are biased to remember norm-violators better than those who abide by the rules of cooperation. Where exactly this bias comes from; selective attention, better storage or retrieval from memory remains an open question.

Finally, in Chapter 7 I investigated a possible link between Theory of Mind (ToM) capacity and the ability to recognise cooperative intentions from faces. In contrast to the intuitive hypothesis that these two skills would be related mainly because of the emotional component involved in both interpreting others' mental states and cooperative intentions, I found no evidence for a relationship between the two. Moreover, my results did not support previous reports suggesting high human proficiency in reading cooperative intentions from immediately available cues. A possible explanation for my findings is the evolutionary arms race between accurately guessing and effectively masking cooperative intentions. It is unlikely that emotions not relating to cooperation such as happiness, surprise or fear had to be masked frequently in the human evolutionary past and that such masking would yield considerable benefits. In contrast, successful masking of emotions relating to cooperation might have led to more profitable interactions. Even if interpreting cooperative intentions somehow associates with ToM, the pressure to mask cues of deception might have affected such a relationship. In the Discussion I stressed the importance of disseminating non-significant results in order to avoid the file drawer problem.

In conclusion, the findings of this thesis provide interesting insights on human cooperation. The main advantage of research presented here is that it offers a new picture of human cooperative behaviour of high explanatory power. My framework incorporates knowledge from different disciplines such as evolutionary psychology, social psychology, anthropology and behavioural economics. The results I obtained encourage treating reputation building behaviour as a robust mechanism shaping human cooperation.

8.2. Limitations and future directions

The results of my thesis suffer from several limitations. Most importantly, all experiments have been conducted with samples taken from student populations. According to Henrich et al. (2010) there may be dramatic behavioural differences between people from various societies. Because research on CA has started developing only recently, it is justified to test initial hypotheses using student participants (Gächter, 2010). To my knowledge there exists only one report of CA in the field. Macfarlan (2010) described labour exchange driven by CA occurring in a Dominican village. Labour spent on bay oil distillation predicted altruistic reputations. The larger the group size and in consequence the competition the more labour people offered which was interpreted as calibrating the cost of a cooperative signal based on its value (Macfarlan, 2010). Moreover, altruistic reputations predicted the number of reciprocal partnerships men formed, hence the benefits of CA translated not necessarily to the quality but the quantity of social partners. Reports like Macfarlan's (2010) deepen the experimental and theoretical knowledge and verify whether the proposed hypotheses can be applied to real-life situations. Despite low external validity of my results, they capture some interesting mechanisms of reputation building under controlled conditions which would be difficult to observe outside the lab.

The studies on CA described in this thesis would benefit from supporting the hypotheses with theoretical models. Hypotheses presented here were derived from the assumptions of CA, however, it has not been validated whether they are theoretically sound. For example, when considering the cost of cooperation, sending a costly signal may not always be the best strategy (Maynard Smith and Harper, 2003). Individuals should seek the ways of signalling that are honest and reliable enough to function as an advertisement but not more costly than that. If a signal has a quantitative nature (as it is with contributions) there may be some optimal contribution that informs participants about the cooperativeness of the signaller. One does not necessarily have to spend all of what they have in order to attract others; some smaller amount may be sufficient and less costly to the signaller. A theoretical model would help to investigate whether people optimize their costly investments. Another problem that requires theoretical approach is the evaluation of the two reputation-based mechanisms of cooperation, CA and IR. My research suggests that CA affects cooperation in social dilemmas to a greater extent than IR. However, it would be beneficial to have a model contrasting how these two work. Simulations would help to determine whether CA or IR is more likely to evolve and be more stable.

The study described in Chapter 4 could benefit from another control. The relationship between heterogeneity in resources and reputation-based cooperation can be tackled from two perspectives. Either the CA setting affects how people behave in a social dilemma situation with heterogeneous resources, or the inequality in resources (e.g. the fact of having £10 rather than £20) affects individuals' reputation building behaviour. In the case of the former assumption, in order to test whether the behaviour of individuals with different resources is a result of the reputation building setting, it is necessary to compare it to a condition with an anonymous setting (e.g. studies cited in Chapter 4 such as Cress and Kimmerle, 2008). If we assume an inverse mechanism of causation, namely, that resource inequality affects reputation building behaviour, a control would entail three types of groups (£10, £15 and £20) with homogenous resources playing a social dilemma game in a reputation building setting. Including the latter control would eliminate the possibility that the high relative contributions of low-resource individuals observed in my study were not due to the willingness to outperform others, but were simply a result of being endowed with a higher or lower amount of money. Although unlikely, it might be argued that three low-resource individuals playing a PGG with each other would contribute proportionately more than three high-resource individuals. The extra control could be implemented to clarify that this is not the case.

The results of the study described in Chapter 6 would be more reliable if there was another person naïve to the hypothesis assessing the information recalled by participants. My assessments and the assessments of the other person should be congruent in order to eliminate any experimental bias. If further resources were available I would spend them on employing a second assessor and making the procedure more rigorous. I would also be tempted to complement Chapter 7 with another experiment investigating personality characteristics of people who can effectively mask their cooperative intentions. Hypothetically, such people would have a higher incentive to cheat and might exhibit higher Machiavellian intelligence than those whose intentions are easily interpreted.

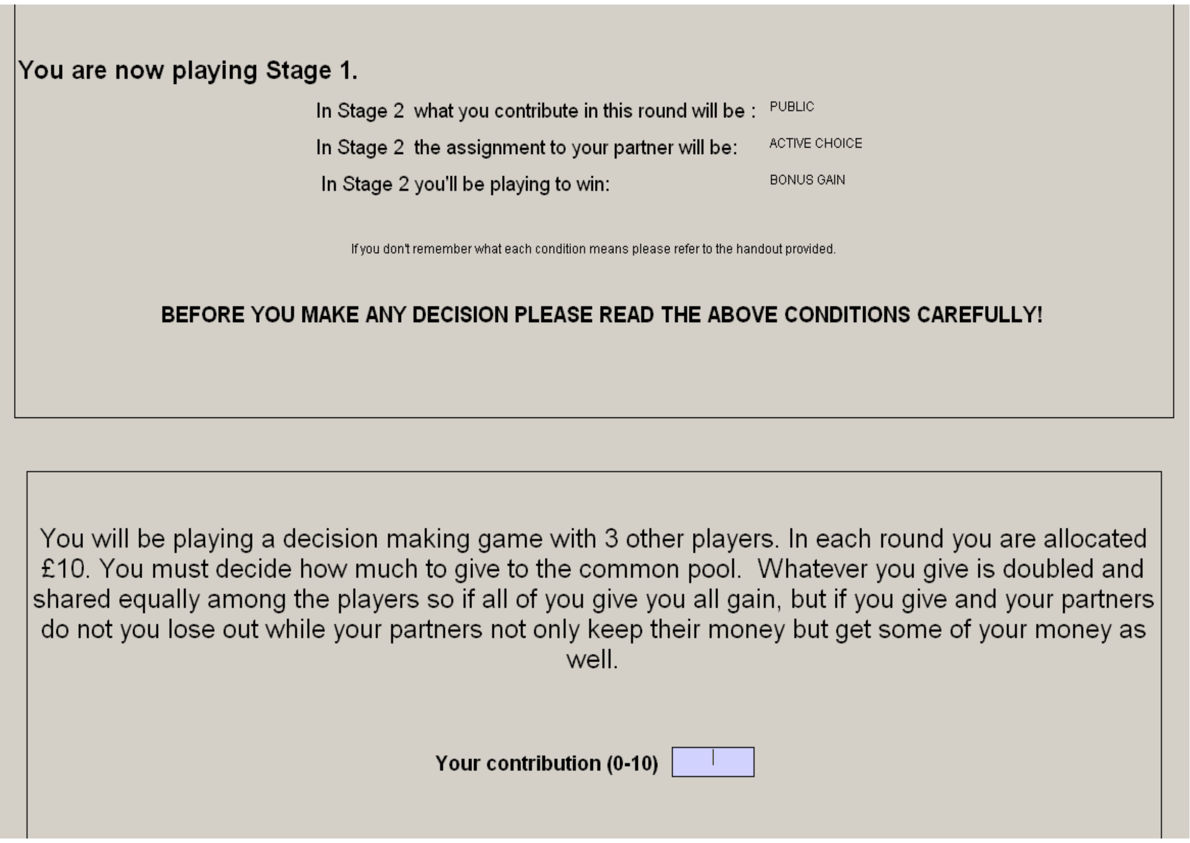
Outperforming others in cooperation when reputation is at stake constitutes the central notion of CA. However, it has not yet been directly tested whether such altruistic competition exists. The experiment described in Chapter 4 provided indirect support for this notion, but in order to unequivocally verify whether people strive to be more cooperative *than others* in contrast to being simply more cooperative in general I propose the following design. Participants play a repeated PGG but in each round the contribution decision of one randomly selected participant is delayed. While others make a simultaneous decision, one person is able to see their contributions before he/she decides how much to contribute. The design involves different conditions e.g. anonymous, knowledge and partner choice (compare to Chapter 2) so it is possible to observe in what way the randomly chosen players adjust their cooperation level to others' contributions. Alternatively, all participants contribute simultaneously, but before they make their decision, they are asked to indicate the average amount of money they expect the others in the group would contribute. If the amount of money contributed is higher than the observed contributions in the first proposed design or the average expected contributions in the second design, it is an evidence for reputational competition. Such experiments would unambiguously determine whether people try to outdo others in cooperative displays.

Finally, an intriguing research topic is the relative importance of reward, punishment and reputation in human interactions. Previous studies have shown that reward (Rand et al., 2009), punishment (Fehr and Gächter, 2002) and reputation (Milinski et al., 2002b; see also Chapter 5) can all enhance cooperation. It would be interesting to see which of the three mechanisms people are most willing to use in order to promote cooperation. Essentially, will people be willing to punish and reward in a situation when cooperation is driven by reputations that bring long-term profits? My guess is that cooperation based on reputation would be sufficient to discourage free-riding and there would be no need to resort to positive or negative sanctions. A recent report indicates that altruistic punishment does not solve the 'tragedy of the commons' in some countries (Herrmann et al., 2008) suggesting that another, more robust mechanism is required to explain large-scale cooperation. I consider reputation building as a promising candidate.

In sum, my thesis answers some novel research questions pertaining to reputation-based cooperation but it also points to challenging research directions and calls for a cross-cultural study of reputation building behaviour.

Appendix A

An example of instructions participants could see in the screen in the study described in Chapter 3.



You are now playing Stage 1.

In Stage 2 what you contribute in this round will be : PUBLIC

In Stage 2 the assignment to your partner will be: ACTIVE CHOICE

In Stage 2 you'll be playing to win: BONUS GAIN

If you don't remember what each condition means please refer to the handout provided.

BEFORE YOU MAKE ANY DECISION PLEASE READ THE ABOVE CONDITIONS CAREFULLY!

You will be playing a decision making game with 3 other players. In each round you are allocated £10. You must decide how much to give to the common pool. Whatever you give is doubled and shared equally among the players so if all of you give you all gain, but if you give and your partners do not you lose out while your partners not only keep their money but get some of your money as well.

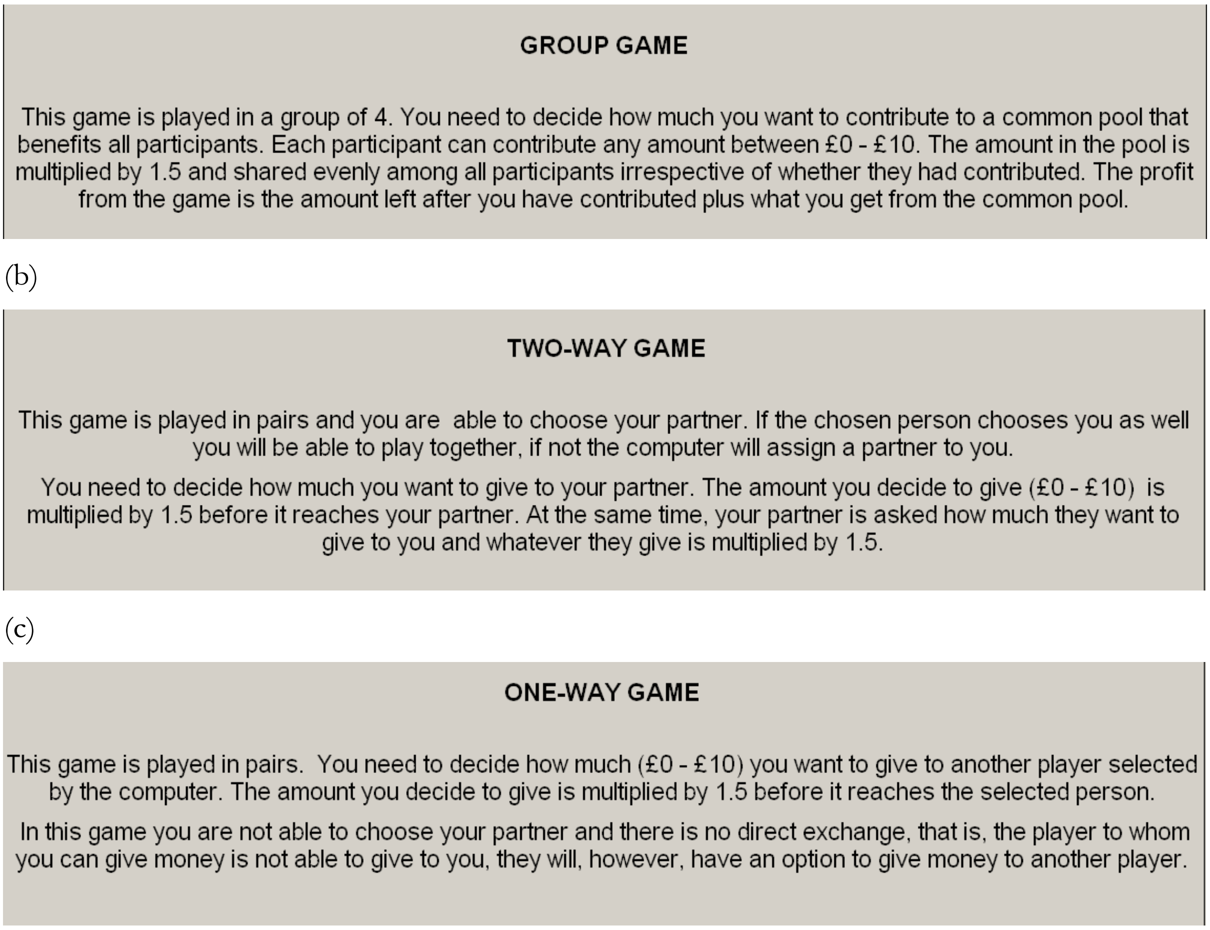
Your contribution (0-10)

Appendix B

Details of the procedure

Participants did the experiment in groups of four; there were six two-male-two-female groups, three all-female groups, six three-male-one-female groups and five three-female-one-male groups. Participants started the experiment by filling out a demographics questionnaire and reading about the three games they would play: group game, one-way game and two-way game. Participants learned that they would have a starting account of 300 lab pounds (£6 real pounds). For each round participants would be able to spend between £0-10. As there were 30 rounds in each experimental session (which participants did not know) it was not possible to go bankrupt before the end. Participants were told that the experiment would take between 30 and 45 minutes. Participants read the instructions for each game and answered test questions (see Figure 1).

Figure 1 Instructions for the three games used in the experiment



(a)

GROUP GAME

This game is played in a group of 4. You need to decide how much you want to contribute to a common pool that benefits all participants. Each participant can contribute any amount between £0 - £10. The amount in the pool is multiplied by 1.5 and shared evenly among all participants irrespective of whether they had contributed. The profit from the game is the amount left after you have contributed plus what you get from the common pool.

(b)

TWO-WAY GAME

This game is played in pairs and you are able to choose your partner. If the chosen person chooses you as well you will be able to play together, if not the computer will assign a partner to you.

You need to decide how much you want to give to your partner. The amount you decide to give (£0 - £10) is multiplied by 1.5 before it reaches your partner. At the same time, your partner is asked how much they want to give to you and whatever they give is multiplied by 1.5.

(c)

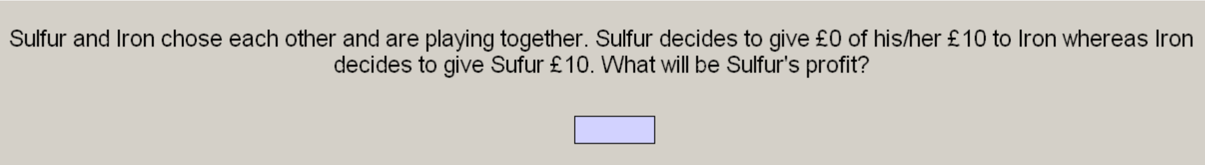
ONE-WAY GAME

This game is played in pairs. You need to decide how much (£0 - £10) you want to give to another player selected by the computer. The amount you decide to give is multiplied by 1.5 before it reaches the selected person.

In this game you are not able to choose your partner and there is no direct exchange, that is, the player to whom you can give money is not able to give to you, they will, however, have an option to give money to another player.

To ensure that participants understood the instructions they had to read, scenarios were presented in which players contributed/donated different amounts of money and participants then had to calculate their profits. For each game participants were presented with two different scenarios, e.g. in the group game a scenario describing a very cooperative group and one describing a group with free-riders. For the one- and two-way games participants had to answer questions about the profits of two players who transferred the same amount of money to others and two players with unequal payoffs (see Figure 2). The order of scenarios was counter-balanced across the games. Participants could use a calculator provided by the software Z-Tree to calculate the outcomes of the hypothetical games. If they inserted an incorrect response a message appeared on the screen informing them about the mistake and reminding them about the rules of the game and how the payoffs should be calculated. In the very rare cases when participants kept inserting an incorrect response one of the researchers (K.S.) approached them and explained the problem. After the training stage participants had an opportunity to raise a hand and ask a question if anything was still unclear to them but no one ever used this opportunity.

Figure 2 An example of a test question for the two-way game



Sulfur and Iron chose each other and are playing together. Sulfur decides to give £0 of his/her £10 to Iron whereas Iron decides to give Sufur £10. What will be Sulfur's profit?

Next, depending on the condition, participants were told that they would play (a) several rounds of the group game, (b) several rounds of the group game each alternated with rounds of the two-way game, (c) several rounds of the group game each alternated with rounds of the one-way game, (d) several rounds of the group game each followed by four rounds of the one-way game, (e) several rounds of the group game each followed by four rounds of the two-way game. Hence, participants did not know how many rounds of a certain combination of games they would play.

Appendix C

Examples of conversation scripts used in the study described in Chapter 6.

a) NEGATIVE REPUTATION

A: Have I told you about Jenny?
B: No, what did she do?
A: We went to the restaurant yesterday and everything was fine until the waiter turned up with a bill. She excused herself saying she has to go to the bathroom and disappeared. Eventually I paid for her dinner.

b) POSITIVE REPUTATION

A: Have you heard about Alex?

B: No, what's new about him?
A: I met him in a bank this week. He donated his earnings from the last month to a charity organisation. He wants to help poor children. He said he would like to establish a charity organisation by himself.

c) NO REPUTATION SOCIAL

A: You know that Martin's sick.
B: Yeah, I heard about it. What is it exactly?
A: I visited him in the hospital yesterday. He said he simply lost consciousness when he was walking back home. They are doing some tests to find out the reason but the doctor could not tell me anything.

d) NON-SOCIAL

A: And? Do you like it?
B: It's delicious! Is it difficult to make?
A: You simply take eggs and separate the yolks from whites. Then you mix the yolks with flour, sugar, and ginger. Preheat the oven and bake it for 15 minutes. I also add some powdered sugar at the end.

Appendix D

An example of photographs used in the study described in Chapter 7.

A person who defected in the PD game.

A person who cooperated in the PD game.

REFERENCES

AKTIPIS, C. A. 2004. Know when to walk away: Contingent movement and the evolution of cooperation. *Journal of Theoretical Biology,* 231, 249-260.

ALEXANDER, R. D. 1987. *The Biology of Moral Systems,* New York, Aldine de Gruyter.

ALI, F. & CHAMORRO-PREMUZIC, T. 2010. Investigating Theory of Mind deficits in nonclinical psychopathy and Machiavellianism. *Personality and Individual Differences,* 49, 169-174.

ANDREONI, J. 1995. Cooperation in public-goods experiments: Kindness or confusion? *American Economic Review,* 85, 891-904.

ANDREONI, J. & PETRIE, R. 2004. Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of Public Economics,* 88, 1605-1623.

ANDREONI, J. & VESTERLUND, L. 2001. Which is the fair sex? Gender differences in altruism. *Quarterly Journal of Economics,* 116, 293-312.

ARONSON, E., WILSON, T. D. & AKERT, R. M. 2004. *Social Psychology,* Upper Saddle River, NJ, Prentice Hall.

BARCLAY, P. 2004. Trustworthiness and competitive altruism can also solve the "tragedy of the commons". *Evolution and Human Behavior,* 25, 209-220.

BARCLAY, P. 2008. Enhanced recognition of defectors depends on their rarity. *Cognition,* 107, 817-28.

BARCLAY, P. 2010. Competitive helping invades defection and increases with the size of biological markets (unpublished manuscript).

BARCLAY, P. & WILLER, R. 2007. Partner choice creates competitive altruism in humans. *Proceedings of the Royal Society B: Biological Sciences,* 274, 749-753.

BARON-COHEN, S., WHEELWRIGHT, S., HILL, J., RASTE, Y. & PLUMB, I. 2001. The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines,* 42, 241-251.

BATESON, M., NETTLE, D. & ROBERTS, G. 2006. Cues of being watched enhance cooperation in a real-world setting. *Biology Letters,* 2, 412-414.

BERECZKEI, T., BIRKAS, B. & KEREKES, Z. 2007. Public charity offer as a proximate factor of evolved reputation-building strategy: an experimental analysis of a real-life situation. *Evolution and Human Behavior,* 28, 277-284.

BERGSTROM, C. T. & LACHMANN, M. 1997. Signalling among relatives. I. Is costly signalling too costly? *Philosophical Transactions of the Royal Society B: Biological Sciences,* 352, 609-617.

BICCHIERI, C. & LEV-ON, A. 2007. Computer-mediated communication and cooperation in social dilemmas: an experimental analysis. *Politics Philosophy Economics,* 6, 139-168.

BLIEGE BIRD, R., SMITH, E. A. & BIRD, D. W. 2001. The hunting handicap: Costly signaling in human foraging strategies. *Behavioral Ecology and Sociobiology,* 50, 9-19.

BOONE, T. R. & BUCK, R. 2003. Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior,* 27, 163-182.

BOYD, R. & RICHERSON, P. J. 1989. The evolution of indirect reciprocity. *Social Networks,* 11, 213-236.

BROWN, W. M., CONSEDINE, N. S. & MAGAI, C. 2005. Altruism relates to health in an ethnically diverse sample of older adults. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences,* 60, 143-152.

BROWN, W. M., CRONK, L., GROCHOW, K., JACOBSON, A., LIU, K. C., POPOVIC, Z. & TRIVERS, R. 2005. Dance reveals symmetry especially in young men. *Nature*, 438, 1148-1150.

BROWN, W. M. & MOORE, C. 2000. Is prospective altruist-detection an evolved solution to the adaptive problem of subtle cheating in cooperative ventures? Supportive evidence using the Wason selection task. *Evolution and Human Behavior,* 21, 25-37.

BROWN, W. M. & MOORE, C. 2002. Smile asymmetries and reputation as reliable indiicators of likelihood to cooperate: An evolutionary analysis. *In:* SHOHOV, S. P. (ed.) *Advances in Psychology Research.* New York: Nova Science Publishers.

BROWN, W. M., PALAMETA, B. & MOORE, C. 2003. Are there nonverbal cues to commitment? An exploratory study using the zero-acquaintance video presentation paradigm. *Evolutionary Psychology,* 1, 42-69.

BSHARY, R. 2002. Biting cleaner fish use altruism to deceive image-scoring client reef fish. *Proceedings of the Royal Society B: Biological Sciences,* 269, 2087-2093.

BSHARY, R. & GRUTTER, A. S. 2006. Image scoring and cooperation in a cleaner fish mutualism. *Nature,* 441, 975-978.

BURNHAM, T. C. & HARE, B. 2007. Engineering human cooperation : Does involuntary neural activation increase public goods contributions? *Human Nature,* 18, 88-108.

BYRNE, R. W. 1996. Machiavellian intelligence. *Evolutionary Anthropology,* 5, 172-180.

CADSBY, C. B. & MAYNES, E. 1999. Voluntary provision of threshold public goods with continuous contributions: Experimental evidence. *Journal of Public Economics,* 71, 53-73.

CAMERER, C. F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction,* New York, Princeton University Press.

CHAMP, P. A. & BISHOP, R. C. 2001. Donation payment mechanisms and contingent valuation: An empirical study of hypothetical bias. *Environmental and Resource Economics,* 19, 383-402.

CHARNESS, G. & GNEEZY, U. 2008. What's in a name? Anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior and Organization,* 68, 29-35.

CHERULNIK, P. D., WAY, J. H., AMES, S. & HUTTO, D. B. 1981. Impressions of high and low Machiavellian men. *Journal of Personality,* 49*,* 388–400.

CHEVALIER, J. A. & MAYZLIN, D. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research,* 43, 345-354.

CHIANG, Y. S. 2010. Self-interested partner selection can lead to the emergence of fairness. *Evolution and Human Behavior,* 31, 265-270.

CHIAPPE, D., BROWN, A., DOW, B., KOONTZ, J., M., R. & MCCULLOCH, K. 2004. Cheaters are looked at longer and remembered better than cooperators in social exchange situations. *Evolutionary Psychology,* 2, 108-120.

CHRISTOPHER, A. N. & SCHLENKER, B. R. 2000. The impact of perceived material wealth and perceiver personality on first impressions. *Journal of Economic Psychology,* 21, 1-19.

CHRISTOPHER, A. N., WESTERHOF, D. L. & MAREK, P. 2005. Affluence cues and perceptions of helping. *North American Journal of Psychology,* 7, 229-238.

COSMIDES, L. & TOOBY, J. 1992. Cognitive adaptations for social exchange. *In:* BARKOW, J., COSMIDES, L. & TOOBY, J. (eds.) *The Adapted Mind.* New York: Oxford University Press.

CRESS, U. & KIMMERLE, J. 2008. Endowment heterogeneity and identifiability in the information-exchange dilemma. *Computers in Human Behavior,* 24, 862-874.

DARWIN, C. 1859. *The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life,* London, John Murray.

DARWIN, C. 1871. *The Descent of Man, and Selectin in Relation to Sex,* New York, Appleton and Company.

DAWKINS, R. & KREBS, J. R. 1979. Arms races between and within species. *Proceedings of the Royal Society B: Biological Sciences,* 205, 489-511.

DE WAAL, F. B. M. 2003. On the possibility of animal empathy. *In:* MANSTEAD, T., FRIJDA, N. & FISCHE, A. (eds.) *Feelings & Emotions: The Amsterdam Symposium.* Cambridge: Cambridge University Press.

DE WAAL, F. B. M. 2008. Putting the Altruism Back into Altruism: The Evolution of Empathy. *Annual Review of Psychology,* 59, 279-300.

DUNBAR, R. I. M. 1993. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences,* 16, 681-735.

DUNBAR, R. I. M. 1996. *Grooming, Gossip, and the Evolution of Language,* Cambridge, Harvard University Press.

DUNBAR, R. I. M. 1998. The social brain hypothesis. *Evolutionary Anthropology,* 6, 178-190.

DUNBAR, R. I. M. 2003. The social brain: Mind, language, and society in evolutionary perspective *Annual Review of Anthropology* 32, 163-181.

DUNBAR, R. I. M. 2004. Gossip in evolutionary perspective. *Review of General Psychology,* 8, 100-110.

DUNBAR, R. I. M., MARRIOTT, A. & DUNCAN, N. D. C. 1997. Human conversational behavior. *Human Nature,* 8, 231-246.

ECKEL, C. C. & GROSSMAN, P. J. 1998. Are women less selfish than men?: Evidence from dictator experiments. *Economic Journal,* 108, 726-735.

EKMAN, P. & FRIESEN, W. V. 1969. Nonverbal leakage and clues to deception. *Psychiatry,* 32, 88-106.

EMLER, N. 1990. A social psychology of reputation. *European Review of Social Psychology,* 1, 171-193.

ENGELMANN, D. & FISCHBACHER, U. 2009. Indirect reciprocity and strategic reputation building in an experimental helping game. *Games and Economic Behavior,* 67, 399-407.

ENQUIST, M. & LEIMAR, O. 1993. The evolution of cooperation in mobile organisms. *Animal Behaviour,* 45, 747-757.

FARRELLY, D., LAZARUS, J. & ROBERTS, G. 2007. Altruists attract. *Evolutionary Psychology,* 5, 313-329.

FEHR, E. & FISCHBACHER, U. 2003. The nature of human altruism. *Nature,* 425, 785-91.

FEHR, E. & GÄCHTER, S. 2002. Altruistic punishment in humans. *Nature,* 415, 137-40.

FEHR, E. & SCHMIDT, K. M. 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics,* 114, 817-868.

FELISBERTI, F. M. & PAVEY, L. (2010) Contextual modulation of biases in face recognition. *PLoS ONE*, 5(9).

FETCHENHAUER, D., GROOTHUIS, T. & PRADEL, J. 2010. Not only states but traits - Humans can identify permanent altruistic dispositions in 20 s. *Evolution and Human Behavior,* 31, 80-86.

FISCHBACHER, U. 2007. Z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics,* 10, 171-178.

FISCHBACHER, U., GÄCHTER, S. & FEHR, E. 2001. Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters,* 71, 397-404.

FISHER, R. A. 1930. *The Genetical Theory of Natural Selection,* Oxford, Clarendon Press.

FRANK, R. H. 1988. *Passions within Reason,* New York, London, W.W. Norton & Company.

FRANK, R. H., GILOVICH, T. & REGAN, D. T. 1993. The evolution of one-shot cooperation: An experiment. *Ethology and Sociobiology,* 14, 247-256.

GÄCHTER, S. 2010. (Dis)advantages of student subjects: What is your research question? *Behavioral and Brain Sciences,* 33, 92-93.

GERT, B. 2008. *The Definition of Morality* [Online]. Available: <http://plato.stanford.edu/archives/fall2008/entries/morality-definition/>

GINTIS, H., SMITH, E. A. & BOWLES, S. 2001. Costly signaling and cooperation. *Journal of Theoretical Biology,* 213, 103-119.

GOFFMAN, E. 1959. *The Presentation of Self in Everyday Life,* London, Allen Lane The Penguin Press.

GRAZIANO, W. G. & EISENBERG, N. H. 1997. Agreeableness: A dimension of personality. *In:* HOGAN, R., JOHNSON, J. & BRIGGS, S. (eds.) *Handbook of personality psychology.* San Diego: Academic Press.

GÜRERK, Ö., IRLENBUSCH, B. & ROCKENBACH, B. 2006. The competitive advantage of sanctioning institutions. *Science,* 312, 108-111.

HALEY, K. J. & FESSLER, D. M. T. 2005. Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior,* 26, 245-256.

HAMILTON, W. D. 1964. The genetical evolution of social behaviour. I. *Journal of Theoretical Biology,* 7, 1-16.

HAMILTON, W. D. 1975. Innate social aptitudes of man: an approach from evolutionary genetics. *In:* FOX, R. (ed.) *Biosocial Anthropology.* New York: Wiley.

HANLEY, J., ORBELL, J. & MORIKAWA, T. 2003. Conflict, Interpersonal Assessment, and the Evolution of Cooperation: Simulation Results. *In:* OSTROM, E. & WALKER, J. (eds.) *Trust and Reciprocity:Interdisciplinary Lessons from Experimental Research.* New York: Sage.

HARDIN, G. 1968. The tragedy of the commons. *Science,* 162, 1243-1248.

HARDY, C. L. & VAN VUGT, M. 2006. Nice guys finish first: The competitive altruism hypothesis. *Personality and Social Psychology Bulletin,* 32, 1402-1413.

HENRICH, J. 2001. Challenges for everyone: Real people, deception, one-shot games, social learning, and computers. *Behavioral and Brain Sciences,* 24, 414-415.

HENRICH, J., BOYD, R., BOWLES, S., CAMERER, C., FEHR, E. & GINTIS, H. 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies,* Oxford, Oxford University Press.

HENRICH, J., HEINE, S. J. & NORENZAYAN, A. 2010. The weirdest people in the world? *Behavioral and Brain Sciences,* 33, 61-83.

HERRMANN, B., THÖNI, C. & GÄCHTER, S. 2008. Antisocial punishment across societies. *Science,* 319, 1362-1367.

HERTWIG, R. & ORTMANN, A. 2001. Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences,* 24, 383-403.

HERTWIG, R. & ORTMANN, A. 2008. Deception in experiments: Revisiting the arguments in its defense. *Ethics and Behavior,* 18, 59-92.

HESS, N. H. & HAGEN, E. H. 2006. Psychological adaptations for assessing gossip veracity. *Human Nature,* 17, 337-354.

HILL, K., BARTON, M. & HURTADO, M. A. 2009. The emergence of human uniqueness: Characters underlying behavioral modernity. *Evolutionary Anthropology,* 18, 187-200.

HOFMEYR, A., BURNS, J. & VISSER, M. 2007. Income inequality, reciprocity and public good provision: An experimental analysis. *South African Journal of Economics,* 75, 508-520.

HOWELL, D. C. 2002. *Statistical Methods for Psychology,* Belmont, CA, Wadsworth Publishing Co Inc.

HYDE, J. S. 2005. The gender similarities hypothesis. *American Psychologist,* 60, 581-592.

IREDALE, W., VAN VUGT, M. & DUNBAR, R. 2008. Showing off in humans: Male generosity as a mating signal. *Evolutionary Psychology,* 6, 386–392.

ISAAC, R. M., WALKER, J. M. & WILLIAMS, A. W. 1994. Group size and the voluntary provision of public goods. Experimental evidence utilizing large groups. *Journal of Public Economics,* 54, 1-36.

JENSEN, C., FARNHAM, S. D., DRUCKER, S. M. & KOLLOCK, P. Year. Effect of communication modality on cooperation in online environments. *In:* Conference on Human Factors in Computing Systems, 2000 The Hague, The Netherlands. New York: Association for Computing Machinery, 470-477.

JONAITIS, A. 1991. *Chiefly Feasts: The Enduring Kwakiutl Potlatch,* Seattle, University of Washington Press.

JONES, G. 2008. Are smarter groups more cooperative? Evidence from prisoner's dilemma experiments, 1959-2003. *Journal of Economic Behavior & Organization,* 68, 489-97.

KEPPEL, G. & WICKENS, T. D. 2004. *Design and Analysis: A Researcher's Handbook*, New Jersey, Prentice Hall.

KILLINGBACK, T., DOEBELI, M. & KNOWLTON, N. 1999. Variable investment, the Continuous Prisoner's Dilemma, and the origin of cooperation. *Proceedings of the Royal Society B: Biological Sciences,* 266, 1723-1728.

KINTSCH, W. 1974. *The representation of meaning in memory,* Oxford, Erlbaum.

KIYONARI, T. & BARCLAY, P. 2008. Cooperation in social dilemmas: Free riding may be thwarted by second-order reward rather than by punishment. *Journal of Personality and Social Psychology,* 95, 826-842.

KÜMMERLI, R., BURTON-CHELLEW, M. N., ROSS-GILLESPIE, A. & WEST, S. A. 2010. Resistance to extreme strategies, rather than prosocial preferences, can explain human cooperation in public goods games. *Proceedings of the National Academy of Sciences of the United States of America,* 107, 10125-10130.

KURZBAN, R. & HOUSER, D. 2005. Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences of the United States of America,* 102, 1803-1807.

LAMBA, S. & MACE, R. 2010. People recognise when they are really anonymous in an economic game. *Evolution and Human Behavior,* 31, 271-278.

LEDYARD, J. O. 1995. Public goods: a survey of experimental research. *In:* KAGEL J.H, R. A. E. (ed.) *The Handbook of Experimental Economics.* New York: Princeton University Press.

LEIMAR, O. & HAMMERSTEIN, P. 2001. Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society B: Biological Sciences,* 268, 745-753.

LITTLE, A. C., BURT, D. M., PENTON-VOAK, I. S. & PERRETT, D. I. 2001. Self-perceived attractiveness influences human female preferences for sexual dimorphism and symmetry in male faces. *Proceedings of the Royal Society B: Biological Sciences,* 268, 39-44.

LOTEM, A., FISHMAN, M. A. & STONE, L. 2002. From reciprocity to unconditional altruism through signaling benefits. *Proceedings of Royal Society London B,* 270, 199-105.

LYONS, M., CALDWELL, T. & SHULTZ, S. 2010. Mind reading and manipulation:is Machiavellianism related to theory of Mind? *Journal of Evolutionary Psychology,* 8, 261–274.

MACFARLAN, S. J. 2010. *The Logic of Labor Exchange in a Dominican Village: Competitive Altruism, Biological Markets, and the Nexus of Male Social Relations.* PhD thesis, Washington State University.

MAYNARD SMITH, J. & HARPER, D. 2003. Animal Signals. Oxford: Oxford University Press.

MCNAMARA, J. M., BARTA, Z., FROMHAGE, L. & HOUSTON, A. I. 2008. The coevolution of choosiness and cooperation. *Nature,* 451, 189-192.

MEALEY, L. 1996. Enhanced memory for faces of cheaters. *Ethology and Sociobiology,* 17, 119-128.

MESOUDI, A., WHITEN, A. & DUNBAR, R. 2006. A bias for social information in human cultural transmission. *British Journal of Psychology,* 97, 405-431.

MIFUNE, N., HASHIMOTO, H. & YAMAGISHI, T. 2010. Altruism toward in-group members as a reputation mechanism. *Evolution and Human Behavior,* 31, 109-117.

MILGRAM, S. 1963. Behavioral study of obedience. *Journal of Abnormal and Social Psychology,* 67, 371-378.

MILINSKI, M., SEMMANN, D., BAKKER, T. C. M. & KRAMBECK, H. J. 2001. Cooperation through indirect reciprocity: Image scoring or standing strategy? *Proceedings of the Royal Society B: Biological Sciences,* 268, 2495-2501.

MILINSKI, M., SEMMANN, D. & KRAMBECK, H. J. 2002a. Donors to charity gain in both indirect reciprocity and political reputation. *Proceedings of the Royal Society B: Biological Sciences,* 269, 881-883.

MILINSKI, M., SEMMANN, D. & KRAMBECK, H. J. 2002b. Reputation helps solve the 'tragedy of the commons'. *Nature,* 415, 424-426.

MILLET, K. & DEWITTE, S. 2007. Altruistic behavior as a costly signal of general intelligence. *Journal of Research in Personality,* 41, 316-326.

MITANI, Y. & FLORES, N. E. 2009. Demand revelation, hypothetical bias, and threshold public goods provision. *Environmental and Resource Economics,* 44, 231-243.

MOLL, H. & TOMASELLO, M. 2007. Cooperation and human cognition: The Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 362, 639-648.

MØLLERAND, A. P. & JENNIONS, M. D. 2001. Testing and adjusting for publication bias. *Trends in Ecology and Evolution,* 16, 580-586.

NAGANAWA, T., YAMAUCHI, S., YAMAGATA, N., MATSUMOTO-ODA, A. & ODA, R. 2010. Do altruists detect altruists easier than non-altruists? *Letters on Evolutionary Behavioral Science,* 1, 2-5.

NELISSEN, R. M. A. 2008. The price you pay:cost-dependent reputation effects of altruistic punishmant. *Evolution and Human Behavior,* 29, 242-248.

NESSE, R. M. 2007. Runaway social selection for displays of partner value and altruism. *Biological Theory,* 2, 143–155.

NETTLE, D. & LIDDLE, B. 2008. Agreeableness is related to social-cognitive, but not social-perceptual, theory of mind. *European Journal of Personality,* 22, 323-335.

NIKIFORAKIS, N. 2010. Feedback, punishment and cooperation in public good experiments. *Games and Economic Behavior,* 68, 689-702.

NOË, R. & HAMMERSTEIN, P. 1994. Biological markets: Supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behavioral Ecology and Sociobiology,* 35, 1-11.

NOË, R. & HAMMERSTEIN, P. 1995. Biological markets. *Trends in Ecology and Evolution,* 10, 336-339.

NOWAK, M. A. & ROCH, S. 2007. Upstream reciprocity and the evolution of gratitude. *Proceedings of the Royal Society B: Biological Sciences,* 274, 605-609.

NOWAK, M. A. & SIGMUND, K. 1998. Evolution of indirect reciprocity by image scoring. *Nature,* 393, 573-577.

NOWAK, M. A. & SIGMUND, K. 2005. Evolution of indirect reciprocity. *Nature,* 437, 1291-1298.

ODA, R., NAGANAWA, T., YAMAUCHI, S., YAMAGATA, N. & MATSUMOTO-ODA, A. 2009a. Altruists are trusted based on non-verbal cues. *Biology Letters,* 5, 752-754.

ODA, R., YAMAGATA, N., YABIKU, Y. & MATSUMOTO-ODA, A. 2009b. Altruism can be assessed correctly based on impression. *Human Nature,* 20, 331-341.

OHTSUKI, H. & IWASA, Y. 2004. How should we define goodness? - Reputation dynamics in indirect reciprocity. *Journal of Theoretical Biology,* 231, 107-120.

OHTSUKI, H. & IWASA, Y. 2006. The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology,* 239, 435-444.

OHTSUKI, H. & IWASA, Y. 2007. Global analyses of evolutionary dynamics and exhaustive search for social norms that maintain cooperation by reputation. *Journal of Theoretical Biology,* 244, 518-531.

OSTROM, E. 2003. Toward a behavioral theory linking trust, reciprocity, and reputation. *In:* OSTROM, E. & WALKER, J. (eds.) *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research.* New York: Russell Sage Foundation Publications.

PAAL, T. & BERECZKEI, T. 2007. Adult theory of mind, cooperation, Machiavellianism: The effect of mindreading on social relations. *Personality and Individual Differences,* 43, 541-551.

PALAMETA, B. & BROWN, W. M. 1999. Human cooperation is more than by-product mutualism. *Animal Behaviour,* 57, F1-F3.

PANCHANATHAN, K. & BOYD, R. 2003. A tale of two defectors: The importance of standing for evolution of indirect reciprocity. *Journal of Theoretical Biology,* 224, 115-126.

PENNER, L., BRANNICK, M. T., WEBB, S. & CONNELL, P. 2005. Effects on volunteering of the September 11, 2001, attacks: An archival analysis. *Journal of Applied Social Psychology,* 35, 1333-1360.

PFEIFFER, T., RUTTE, C., KILLINGBACK, T., TABORSKY, M. & BONHOEFFER, S. 2005. Evolution of cooperation by generalized reciprocity. *Proceedings of the Royal Society B-Biological Sciences,* 272, 1115-1120.

PIAZZA, J. & BERING, J. M. 2008. Concerns about reputation via gossip promote generous allocations in an economic game. *Evolution and Human Behavior,* 29, 172-178.

POLLOCK, G. & DUGATKIN, L. A. 1992. Reciprocity and the emergence of reputation. *Journal of Theoretical Biology,* 159, 25-37.

PORTER, S. & BRINKE, L. T. 2008. Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological Science,* 19, 508-514.

PRICE, M. E., COSMIDES, L. & TOOBY, J. 2002. Punitive sentiment as an anti-free rider psychological device. *Evolution and Human Behavior,* 23, 203-231.

RAND, D. G., DREBER, A., ELLINGSEN, T., FUDENBERG, D. & NOWAK, M. A. 2009. Positive interactions promote public cooperation. *Science,* 325, 1272-1275.

RANKIN, D. J. & EGGIMANN, F. 2009. The evolution of judgement bias in indirect reciprocity. *Proceedings of the Royal Society B: Biological Sciences,* 276, 1339-1345.

RESNICK, P., ZECKHAUSER, R., SWANSON, J. & LOCKWOOD, K. 2006. The value of reputation on eBay: A controlled experiment. *Experimental Economics,* 9, 79-101.

RHODES, G. 2006. The evolutionary psychology of facial beauty. *Annual Review of Psychology,* 57, 199-226.

RILLING, J. K., GUTMAN, D. A., ZEH, T. R., PAGNONI, G., BERNS, G. S. & KILTS, C. D. 2002. A neural basis for social cooperation. *Neuron,* 35, 395-405.

ROBERTS, G. 1998. Competitive altruism: from reciprocity to the handicap principle. *Proceedings of the Royal Society B: Biological Sciences,* 265, 427-31.

ROBERTS, G. 2008. Evolution of direct and indirect reciprocity. *Proceedings of the Royal Society B: Biological Sciences,* 275, 173-179.

ROBERTS, G. & RENWICK, J. S. 2003. The development of cooperative relationships: An experiment. *Proceedings of the Royal Society B: Biological Sciences,* 270, 2279-2283.

ROBERTS, G. & SHERRATT, T. N. 1998. Development of cooperative relationships through increasing investment. *Nature,* 394, 175-179.

ROSENTHAL, R. 1979. The file drawer problem and tolerance for null results. *Psychological Bulletin,* 86, 638-641.

RUSSELL, Y. I., CALL, J. & DUNBAR, R. I. M. 2008. Image scoring in great apes. *Behavioural Processes,* 78, 108-111.

RUTTE, C. & TABORSKY, M. 2007. Generalized reciprocity in rats. *PLoS Biology,* 5.

SABBAGH, M. A. 2004. Understanding orbitofrontal contributions to theory-of-mind reasoning: Implications for autism. *Brain and Cognition,* 55, 209-219.

SAIJO, T. & NAKAMURA, H. 1995. The spite dilemma in voluntary contribution mechanism experiments. *Journal of Conflict Resolution,* 39, 535–560.

SALLY, D. 1995. Conversation and cooperation in social dilemmas - A metaanalysis of experiments from 1958 to 1992. *Rationality and Society,* 7, 58-92.

SALLY, D. & HILL, E. 2006. The development of interpersonal strategy: Autism, theory-of-mind, cooperation and fairness. *Journal of Economic Psychology,* 27, 73-97.

SCHUG, J., MATSUMOTO, D., HORITA, Y., YAMAGISHI, T. & BONNET, K. 2010. Emotional expressivity as a signal of cooperation. *Evolution and Human Behavior,* 31, 87-94.

SEINEN, I. & SCHRAM, A. 2006. Social status and group norms: Indirect reciprocity in a repeated helping experiment. *European Economic Review,* 50, 581-602.

SEMMANN, D., KRAMBECK, H. J. & MILINSKI, M. 2004. Strategic investment in reputation. *Behavioral Ecology and Sociobiology,* 56, 248-252.

SILK, J. B., BROSNAN, S. F., VONK, J., HENRICH, J., POVINELLI, D. J., RICHARDSON, A. S., LAMBETH, S. P., MASCARO, J. & SCHAPIRO, S. J. 2005. Chimpanzees are indifferent to the welfare of unrelated group members. *Nature,* 437, 1357-1359.

SIMPSON, B. & WILLER, R. 2008. Altruism and indirect reciprocity: The interaction of person and situation in prosocial behavior. *Social Psychology Quarterly,* 71, 37-52.

SMITH, E. A. & BLIEGE BIRD, R. 2005. Costly Signaling and Cooperative Behavior. *In:* GINTIS, H., BOWLES, S., BOYD, R. T. & FEHR, E. (eds.) *Moral Sentiments and Material Interests: The Foundation of Cooperation in Economic Life.* Cambridge: The MIT Press

SMITH, E. A., BLIEGE BIRD, R. & BIRD, D. W. 2003a. The benefits of costly signaling: Meriam turtle hunters. *Behavioral Ecology,* 14, 116-126.

SMITH, E. R., JACKSON, J. W. & SPARKS, C. W. 2003b. Effects of Inequality and Reasons for Inequality on Group Identification and Cooperation in Social Dilemmas. *Group Processes Intergroup Relations,* 6, 201-220.

SOMMERFELD, R. D., KRAMBECK, H. J. & MILINSKI, M. 2008. Multiple gossip statements and their effect on reputation and trustworthiness. *Proceedings of the Royal Society B: Biological Sciences,* 275, 2529-2536.

SOMMERFELD, R. D., KRAMBECK, H. J., SEMMANN, D. & MILINSKI, M. 2007. Gossip as an alternative for direct observation in games of indirect reciprocity. *Proceedings of the National Academy of Sciences of the United States of America,* 104, 17435-17440.

STILLER, J. & DUNBAR, R. I. M. 2007. Perspective-taking and memory capacity predict social network size. *Social Networks,* 29, 93-104.

STIRRAT, M. & PERRETT, D. I. 2010. Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science,* 21, 349-354.

SUBIAUL, F., VONK, J., OKAMOTO-BARTH, S. & BARTH, J. 2008. Do chimpanzees learn reputation by observation? Evidence from direct and indirect experience with generous and selfish strangers. *Animal Cognition*, 1-13.

SUGDEN, R. 1986. *The Economics of Rights, Co-operation and Welfare,* Oxford, Blackwell.

SYLWESTER, K. & ROBERTS, G. 2010. Cooperators benefit through reputation-based partner choice in economic games. *Biology Letters,* 6, 659-662.

TAJFEL, H. 1970. Experiments in intergroup discrimination. *Scientific American,* 223, 96-102.

TAJFEL, H., BILLIG, M., BUNDY, R. & FLAMENT, C. 1971. Social categorization in intergroup behavior. *European Journal of Social Psychology,* 1, 149–78.

TAKAGISHI, H., KAMESHIMA, S., SCHUG, J., KOIZUMI, M. & YAMAGISHI, T. 2010. Theory of mind enhances preference for fairness. *Journal of Experimental Child Psychology,* 105, 130-137.

TODOROV, A., BARON, S. G. & OOSTERHOF, N. N. 2008. Evaluating face trustworthiness: A model based approach. *Social Cognitive and Affective Neuroscience,* 3, 119-127.

TRIVERS, R. L. 1971. The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology,* 46, 35-57.

VAN DEN ASSEM, M. J., VAN DOLDER, D. & THALER, R. H. 2010. Split or Steal? Cooperative Behavior When the Stakes are Large. *Social Science Research Network,* available at SSRN: http://ssrn.com/abstract=1592456.

VAN DIJK, E. & WILKE, H. 1994. Asymmetry of wealth and public good provision. *Social Psychology Quarterly,* 57, 352-359.

VAN DIJK, E. & WILKE, H. 1995. Coordination Rules in Asymmetric Social Dilemmas: A Comparison between Public Good Dilemmas and Resource Dilemmas. *Journal of Experimental Social Psychology,* 31, 1-27.

VAN LANGE, P. A. M., DE BRUIN, E. M. N., OTTEN, W. & JOIREMAN, J. A. 1997. Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology,* 73, 733-746.

VAN VUGT, M. 2001. Self-interest as self-fulfilling prophecy. *Behavioral and Brain Sciences,* 24, 429-430.

VANNESTE, S., VERPLAETSE, J., VAN HIEL, A. & BRAECKMAN, J. 2007. Attention bias toward noncooperative people. A dot probe classification study in cheating detection. *Evolution and Human Behavior,* 28, 272-276.

VEBLEN, T. 1899. *The Theory of the Leisure Class,* New York, Dover Publications Inc.

VERPLAETSE, J., VANNESTE, S. & BRAECKMAN, J. 2007. You can judge a book by its cover: the sequel. A kernel of truth in predictive cheating detection *Evolution and Human Behavior,* 28, 260-71.

WEDEKIND, C. & BRAITHWAITE, V. A. 2002. The long-term benefits of human generosity in indirect reciprocity. *Current Biology,* 12, 1012-1015.

WEDEKIND, C. & MILINSKI, M. 2000. Cooperation through image scoring in humans. *Science,* 288, 850-852.

WEST, S. A., EL MOUDEN, C. & GARDNER, A. (in press). 16 common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior.*

WEST, S. A., GARDNER, A., SHUKER, D. M., REYNOLDS, T., BURTON-CHELLOW, M., SYKES, E. M., GUINNEE, M. A. & GRIFFIN, A. S. 2006. Cooperation and the Scale of Competition in Humans. *Current Biology,* 16, 1103-1106.

WEST, S. A., GRIFFIN, A. S. & GARDNER, A. 2007. Evolutionary Explanations for Cooperation. *Current Biology,* 17, R661 - R672

WHITEN, A. 2000. Social complexity and social intelligence. *In:* BOCK, G., GOODE, J. & WEBB, K. (eds.) *The nature of intelligence.* New York: Cambridge University Press.

WILSON, D. S., NEAR, D. & MILLER, R.R 1996. Machiavellianism: a synthesis of the evolutionary and psychological literatures. *Psychological Bulletin* 119, 285–299.

WILSON, D. S., WILCZYNSKI, C., WELLS, A. & WEISER, L. 2000. Gossip and other aspects of language as group-level adaptations. *In:* HEYES,. C. & HUBER, L. (ed.) *The Evolution of Cognition.* Cambridge, MA: MIT Press.

WILSON, E. O. 1975. *Sociobiology,* Cambridge, Harvard University Press.

WIT, A., WILKE, H. A. M. & OPPEWAL, H. 1992. Fairness in asymmetric social dilemmas. *In:* LIEBRAND, W. B. G., MESSICK, D. M. , WILKE, H. A. M (ed.) *Social Dilemmas: Theoretical Issues and Research Findings* Elmsford: Pergamon Press.

WYER JR, R. S., BUDESHEIM, T. L., LAMBERT, A. J. & SWAN, S. 1994. Person memory and judgment: Pragmatic influences on impressions formed in a social context. *Journal of Personality and Social Psychology,* 66, 254-267.

YAMAGISHI, T. & MIFUNE, N. 2008. Does shared group membership promote altruism?: Fear, greed, and reputation. *Rationality and Society,* 20, 5-30.

YAMAGISHI, T. & MIFUNE, N. 2009. Social exchange and solidarity: in-group love or out-group hate? *Evolution and Human Behavior,* 30, 229-237.

YAMAGISHI, T., TANIDA, S., MASHIMA, R., SHIMOMA, E. & KANAZAWA, S. 2003. You can judge a book by its cover. Evidence that cheaters may look different from cooperators. *Evolution and Human Behavior,* 24, 290-301.

YU, C., AU, W. & CHAN, K. K. 2009. Efficacy = endowment × efficiency: revisiting efficacy and endowment effects in a public goods dilemma. *Journal of Personality and Social Psychology,* 96, 155–169.

ZAHAVI, A. 1995. Altruism as a handicap - The limitations of kin selection and reciprocity. *Journal of Avian Biology,* 26, 1-3.

ZAHAVI, A. & ZAHAVI, A. 1997. *The Handicap Principle: A Missing Piece of Darwin's Puzzle,* New York, Oxford University Press.

ZEBROWITZ, L. A., VOINESCU, L. & COLLINS, M. A. 1996. "Wide-eyed" and "crooked-faced": Determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin,* 22, 1258-1269.