REPUTATIONAL BENEFITS OF ALTRUISM AND ALTRUISTIC PUNISHMENT

REPUTATIONAL BENEFITS OF ALTRUISM AND ALTRUISTIC PUNISHMENT

By
PAT BARCLAY, B.SC.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree
Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2005)
(Psychology)

McMaster University
Hamilton, Ontario

TITLE: Reputational Benefits of Altruism and Altruistic Punishment

AUTHOR: Pat Barclay, B.Sc. (University of Guelph)

SUPERVISORS: Professors Martin Daly and Margo Wilson

NUMBER OF PAGES: xiv, 195

Abstract

Altruism poses a potential problem for evolutionary theory because altruistic individuals provide benefits to others at a cost to themselves, and this cost implies that such behaviour should not evolve. A number of theories have been proposed to account for the existence of apparently altruistic behaviours in nature. Many altruistic acts are directed towards non-kin and do not appear to be reciprocated by others, leading some researchers to propose that cooperative sentiments must have evolved via group selection. However, Zahavi's theory of costly signaling can help explain the evolution of cooperative sentiments, and there has been a recent increase in theoretical and empirical applications of costly signaling theory. When applied to the study of altruism, this theory predicts that altruism can function as an honest signal of unobservable qualities such as abilities, resources, or cooperative intent, so long as the cost of the altruism is sufficiently high to discourage such behaviour in individuals who do not actually possess such qualities. After reviewing the various theories that could potentially account for the evolution of altruism (Chapter 1), I test some predictions about human cooperation derived from costly signaling theory. In Chapter 2, I show that experimental participants were more cooperative when they had cues that they could benefit from having a good reputation, and that there was apparently some competition to be the most generous group member. Furthermore, I show that people tended to trust group members who are cooperative in other contexts (replicated in Chapter 4). Chapter 3 failed to find evidence that artificially

granting high status to people makes them more likely to contribute to public goods or punish free-riders, but there was suggestive evidence that physical proximity to the experimenter affected contributions and punishment. In Chapter 4, I found that people tended to trust others who were willing to incur costs to punish those who free-ride on group cooperation, and that men were more punitive than women. In Chapter 5, I present evidence that women find altruistic men more desirable than neutral men for long-term relationships. Together, these results suggest that humans do treat altruism as a signal of willingness to be cooperative. These findings are discussed with respect to the adaptive design of cooperative sentiments as well as the current debate over group selection.

Acknowledgements

First of all, I would like to thank Martin Daly and Margo Wilson for their knowledge, their advice, their support, their patience, and their food. I have learned much from them about evolution, psychology, experimental design and rigor, the joys of academia in general, and many other things, and without them none of this would have been possible. I would also like to thank Andrew Muller for his comments and advice, and for helping me to learn more about the experimental economics side of my work. I would like to thank other graduate students and post-doctoral researchers of the Daly/Wilson lab, past and present, including Lisa DeBruine, Andrew Clark, Danny Krupp, Greg Dingle, Sean Myles, Toko Kiyonari, Steve Stewart-Williams, Nick Pound, and Catherine Salmon, for good discussions, advice, collaborations, fun, juggling, other assorted lab antics, and some examples to look up to as guidelines on "how to do it right". Paul Ramos deserves much praise and gratitude for designing all of the computer programs used in this thesis, and other research assistants whom I owe thanks to include Angela Chang, Melanie Mackenzie, Abby Morrison, and Agnes Rekas. Other members of the McMaster University Psychology Department who deserve mention include Eric Bressler, Aimee Skye, Rolfe Morrison, all members of the Animal Behaviour Journal Club, fellow graduate students for some great times, the wonderful Psychology Department office staff for their help and support, other professors for good advice and knowledge, and Gary Weatherill for saving my computer from death not just once, but twice.

I would also like to thank my parents, Walter and Judy, for their love and support and for helping me along this far. I obviously wouldn't be here without help from them. I also thank my brother Michael for love and inspiration and being something to live up to. I also want to thank the friends and family that I've enjoyed spending time with, especially those who I have enjoyed having countless discussions with about various aspects of evolution, psychology, or interesting topics in general. You have all helped me stay sane, or at least as sane as could reasonably be expected. They are too many to name individually, but Simon Racioppa deserves special recognition for helping introduce me to critical thinking and for his enthusiasm about my work and evolutionary psychology in general. Thanks also to my undergraduate thesis advisor, Hank Davis from the University of Guelph, for helping me get started on this path. I want to mention the other researchers that got me interested in evolution in general and inspired me to do research, but they are also too numerous to name, and they will know who they are when they see me citing them. Many thanks to the musicians (especially the independent or local ones) who have helped me through many a late night working, but they are definitely far too numerous to name. Finally, I would like to thank and sincerely apologize to anyone whom I should have mentioned by name and did not, and can only plead that my habit of writing things late at night sometimes lead me to miss important details (such as important people), and hope that they forgive me.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 General Introduction

An organism is said to act altruistically when it does something that is costly to itself yet beneficial to another. Evolutionary biologists have sought to explain the existence of altruistic behaviour by considering the selective forces that may have caused the evolution of mental decision-rules that cause altruistic behaviour. This thesis continues that tradition by investigating some possible benefits that altruists might receive for acting altruistically, in order to infer how such behaviour could have evolved. This does not mean to imply that altruistic acts are consciously deliberated, or that people always have ulterior motives and seek to benefit from being nice to others. I am using a definition of altruism that looks only at the costs to the altruist and the benefits to the recipient rather than the particular motivations that underlie such behaviour. People may be legitimately concerned for others and be genuinely motivated to aid others simply because they have the welfare of others as a goal, and these sentiments will cause them to act in a prosocial manner. Also, if a generous act does happen to bring benefits to the altruist at a later point in time, this does not mean the act was not altruistic at the particular time it was performed. Rather than investigating the particular motivational mechanisms that underlie altruistic behaviour, I am investigating why people might have the sentiments that cause such behaviour, and I am examining the cues and incentives that trigger altruism in order to make inferences about what selective forces might have shaped the capacity to develop

such motivations. This question of functional design is separate and complementary to the questions of what the particular psychological motivations or developmental causes are (Tinbergen, 1968). Even if people do have entirely unselfish motives, that would say nothing about the evolution of such motives (Sober & Wilson, 1998). If having cooperative sentiments and acting on them tends to bring benefits to altruists, then such sentiments will have tended to increase in prevalence in populations via biological evolution across generations. This argument does not necessarily imply that people do not learn how much altruism to perform. If acting prosocially brings personal benefits (i.e. it is rewarded), then such behaviour will increase in frequency as individuals learn to behave cooperatively, provided that they already possess an evolved capacity to learn the relationship between altruistic acts and the benefit that they bring. Thus, I am investigating the types of benefits that altruists might receive.

1.2 Why is Altruism a Problem?

For decades, evolutionary biologists have sought to explain the existence of apparently altruistic behaviour in nature. When an organism acts altruistically, it benefits others at a cost to itself. As long as such benefits and costs translate in some way to fitness benefits and costs, then altruists would be at a selective disadvantage. Altruistic organisms would leave fewer offspring, causing a decrease in the proportion of organisms that possess causal mechanisms for such behaviour. Thus, unselfish behaviour would tend to decrease in prevalence in

populations unless other selective pressures counteract that disadvantage. Until the 1960s, many researchers claimed that unselfish behaviour could evolve because it was "good for the species". George Williams (1966) noted that many of the so-called examples of cooperative behaviour were better interpreted as being adaptations to increase individual fitness rather than the fitness of populations or species. He noted that a selfish individual in a group of altruists would have higher fitness than the altruists. Selfishness would then spread through the group and undermine levels of cooperation, making cooperation unlikely to evolve via differential reproduction of groups (which became known as "group selection"). Since approximately that time, many researchers have steered away from group-level explanations and have focused on the individual-level factors that would make certain behaviours or characteristics (and the genes or sets of genes that cause them) increase in prevalence (e.g. Dawkins, 1976; Hamilton, 1964; Maynard Smith & Price, 1973).

For example, Hamilton (1964) realized that acts that appear altruistic from the perspective of the individual may be selfish from the perspective of the gene. He mathematically proved that a gene (or set of genes) could increase in prevalence in a population by benefiting copies of itself present in any individuals sharing a recent common ancestor, such as offspring or close kin. This idea, known as inclusive fitness theory (and later often referred to as kin selection), has had a great impact on evolutionary biology and especially behavioural ecology. Countless studies have investigated the significance of inclusive fitness in non-

human animals (see for a review: Alcock, 1993; Daly & Wilson, 1983; Dugatkin, 2004), and several have focused on humans (e.g. Betzig & Turke, 1986; Daly & Wilson, 1988; DeBruine, 2002; Grayson, 1993; Hames, 1987; Petrinovich, O-Neill, & Jorgensen, 1993). As powerful as this idea is, it is unlikely to explain all altruistic behaviour because many such acts appear to be systematically directed towards non-kin, and the cost of this makes it unlikely that this behaviour is merely a byproduct of mechanisms whose evolved function is nepotistic. This is particularly true in humans, because humans spend significant time and energy cooperating with non-kin.

1.3 Dyadic Relationships and Reciprocal Altruism

1.3.1 Direct Reciprocity

Trivers (1971) introduced the concept of reciprocal altruism, in which individuals who reciprocate altruistic acts towards each other can outcompete others who do not, provided that they can distinguish between others and direct their generosity towards others that have reciprocated in the past. Reciprocal altruists reap the benefits of mutual cooperation, yet do not get taken advantage of by non-cooperators. Using a computer "tournament" of strategies designed to imitate social evolution, Axelrod and Hamilton (1981) provided an early demonstration that a strategy of reciprocal altruism could evolve. Their simulation involved agents playing a two-player cooperative game called the "Prisoner's Dilemma" (PD) in which players have two moves, cooperate or

defect, and the payoffs are structured such that defection is the dominant strategy but mutual cooperation pays better than mutual defection. Thus, each individual has a selfish incentive to defect, but both individuals are worse off if both do so than if both cooperate. Axelrod and Hamilton had a number of computer strategies play a series of iterated PD games with each other, and noted that the most successful strategies started out by cooperating but repaid defection with defection. The most successful strategy was "Tit for Tat", which starts by cooperating and simply imitates the previous move of its partner, providing a classic example of how the capacity for reciprocal altruism can provide a selective advantage.

Much work has since been done using the PD as a model for cooperative interaction. For example, the presence of occasional defectors due to mutation or error allows conditional cooperators (such as Tit for Tat) to dominate unconditional cooperators (McNamara, Barta, & Houston, 2004; Nowak & Sigmund, 1992). Tit for Tat itself can be dominated by strategies that are more forgiving or that will exploit unconditional cooperators (Nowak & Sigmund, 1992, 1993). When agents can vary their cooperation levels continuously instead of discretely, a very successful strategy is to respond to reciprocity by increasing levels of cooperation ("raise the stakes", Roberts & Sherratt, 1998; Sherratt & Roberts, 1999).

Many researchers have claimed to find evidence of reciprocal altruism in non-human animals. For example, vampire bats preferentially regurgitate blood

towards others from whom they have received blood (Wilkinson, 1984), sticklebacks prefer to inspect predators with conspecifics who have previously demonstrated a willingness to approach predators (Milinski, Külling, & Kettler, 1990; Milinski, Pfluger, Külling, & Kettler, 1990), primates tend to groom, support, or give food to others that have done so to them in the past (e.g. Barrett, Henzi, Weingrill, Lycett, & Hill, 2000; Hauser, Chen, Chen, & Chuang, 2003; Watts, 2002), and red-winged blackbirds do not perform as much cooperative nest defence with neighbours who have been prevented from cooperating in the past (Olendorf, Getty, & Scribner, 2004). However, alternative explanations have been advanced for many instances of apparent reciprocity, including confounding reciprocity with kinship (Hammerstein, 2003), and byproduct mutualism (Connor, 1996). Some researchers have explicitly noted a dearth of evidence that strongly supports reciprocal altruism in non-human animals or at least non-primates (Hammerstein, 2003; Noë, 1990), so it is fair to say that the evidence for reciprocal altruism in non-humans is equivocal, or at least is not as widespread as many researchers would like to believe.

The evidence for reciprocal altruism in humans is more straightforward, and some form of reciprocity is present in all human societies (Brown, 1991). Numerous laboratory studies have shown that people behave as if they are concerned with reciprocity (e.g. Berg, Dickhaut, & McCabe, 1995; Cox, 2004; Fehr, Fischbacher & Gächter, 2002; Komorita & Parks, 1995; Roberts & Renwick, 2003). Outside of laboratories, reciprocity seems to be a good

explanation of such diverse phenomena as information sharing among lobster fishermen (Palmer, 1991), food sharing in some (but not all) hunter-gatherer or horticultural tribes (e.g. Dwyer & Minnegal, 1997; Gurven, Hill, Kaplan, Hurtado, & Lyles, 2000; Gurven, Allen-Arave, Hill, & Hurtado, 2001; Patton, 2005), labour exchange (Hames, 1987), restaurant tipping (Strohmetz, Rind, Fisher, & Lynn, 2002), and the "live-and-let-live" policies of soldiers engaged in trench warfare (Axelrod, 1984).

Based on evidence that people are particularly good at solving logic problems that involve detecting instances of social contracts being broken, Cosmides and Tooby (1992) argued that humans have specialized cognitive mechanisms for detecting cheaters in reciprocal relationships in order to avoid being taken advantage of. Mealey, Daood, and Krage (1996) found that people had better memory for the faces of putative low-status cheaters than for other people. There has been considerable debate about the specificity of cognitive mechanisms involved in these cheater-detection and cheater-recognition phenomena and whether they are specifically designed for detecting cheaters (e.g. see Atran, 2001; Barclay & Lalumière, *in press*; Cheng & Holyoak, 1989; Fodor, 2000; Staller, Sloman, & Ben-Zeev; Stone et al., 2002). However, the fact remains that humans are very good at detecting instances of cheating, and humans tend to cooperate much less when faced with non-cooperators (e.g. Fischbacher, Gächter, & Fehr, 2001; Monteresso, Ainslie, Toppi Mullen, & Gault, 2003). Thus, humans may possess cognitive mechanisms that function to support reciprocal

altruism even if those mechanisms also allow humans to solve other problems or evolved for somewhat more general purposes.

1.3.2 Indirect reciprocity

When recipients of altruism can (and do) reciprocate directly to the altruist, this is known as direct reciprocity. Sometimes altruism can be reciprocated indirectly, i.e. by individuals other than the beneficiary of the altruism (Alexander, 1987). In such a system of indirect reciprocity, each individual provides benefits only to those who have done so to others in the past (even if he/she has not received something from them directly), and receive more benefits themselves if they have cooperated in the past. In this way, high levels of cooperation are maintained and non-cooperators are excluded from benefiting. Nowak and Sigmund (1998a, b) provided a mathematical model of indirect reciprocity whereby agents develop a positive reputation for cooperating and only cooperate with others whose score is above a threshold "image score", and showed that this strategy cannot be invaded by defectors. Wedekind and Milinski (2000) had people play an experimental game in which they could donate money to others and were given information about the donating histories of potential recipients. Although participants never interacted with each other twice and had no opportunity to reciprocate generosity directly to benefactors, they tended to give more often to potential recipients who had given to others. Participants who gave the most often tended to receive the most donations. Other researchers using

similar methods have reported similar results (Bolton, Katok, & Ockenfels, *in press*; Seinen & Schram, *in press*).

Other theorists have presented alternative models of indirect reciprocity (Leimar & Hammerstein, 2001; Mashima & Takahashi, 2003; Panchanathan & Boyd, 2003; Takahashi & Mashima, 2004), and have claimed that they are evolutionarily stable under a wider range of conditions than Nowak and Sigmund's (1998a, b) "image scoring" model. Many such models use some form of "standing strategy", whereby agents acquire "good standing" by donating to others and "bad standing" by defecting on cooperators, but remain in good standing if they defect on defectors. In such systems, potential donors of aid give only to those in good standing, and treat defections against non-cooperators as justified defections. The concept of justified defections makes intuitive sense, and this kind of indirect reciprocity prevents cooperators from punishing each other for punishing non-cooperators. There have been attempts to determine which of these strategies provide a more accurate description of what people actually do (Bolton et al., *in press*; Milinski, Semmann, Bakker, & Krambeck, 2001). The results are somewhat mixed, but tentatively imply that people perform indirect reciprocity using decision rules that are more similar to image scoring models than to standing strategies. Despite this debate about the particular form of indirect reciprocity that is most likely to evolve and which form is found in humans, there is general consensus that some forms of indirect reciprocity are

evolutionarily stable and that people do perform some form of indirect reciprocity.

Some field evidence for indirect reciprocity comes from Gurven, Allen-Arave, Hill, and Hurtado (2001), who found that hunters who often shared food tended to receive more food from others when sick and received food from more people, than hunters who could not or would not share as often. This could be characterized as indirect reciprocity, or it could be the outcome of group members following their own self-interest by ensuring the health of good meat-providers. By providing for others, such hunters are making themselves indispensable to the group, and such indispensability gives others an incentive to be altruistic to them to ensure their continued presence in a group (Kaplan & Hill, 1985; Tooby & Cosmides, 1996).

1.4 The Problem of Collective Action

1.4.1 Introduction to Public Goods

Systems of direct and indirect reciprocity both rely on individuals being able to target their altruism specifically towards cooperators while excluding non-cooperators from benefiting. However, there are many situations in which this is not possible, such as the provision of public goods or restraint from overharvesting a common pool resource. A public good is something that people have to incur costs to provide, yet others can benefit from it being provided whether or not they themselves helped to provide it (Davis & Holt, 1993, Messick

& Brewer, 1983), so the public good is vulnerable to exploitation by free-riders. Some examples of public goods for humans include vigilance, group protection, irrigation, and any collective action project. In their simplest form, public goods are comparable to multiple-player Prisoner's Dilemmas. The provision of a public good is collectively beneficial, but free-riders who cooperate relatively little are better off than cooperators who provide the public good, causing selection for non-cooperation that should eventually undermine collective action. Restraint from overharvesting a common pool resource is a public good because overharvesting is individually beneficial but collectively detrimental, such that a "tragedy of the commons" occurs as the resource gets used up and destroyed by each individual following his/her selfish incentive to overharvest (Hardin, 1968). There are slight differences between the provision of public goods and "tragedies of the commons", but they both share the important property that selfish individuals cannot be excluded from benefiting from the cooperation (i.e. provision of the good or restraint from overharvesting) of others.

Although modern society has many public goods that would not have been present in ancestral times (e.g. public television, national defense, scientific research), ancestral humans would have faced many potential public goods situations such as group defense or the policing of group norms. Big-game hunting in many hunter-gatherer societies is a potential public good (Hawkes, 1993) that has received much study. Hunters in some groups focus on big game that can be shared easily and is difficult to acquire, despite being able to earn a

higher private rate of return from other resources that are easier to acquire and less easily shared (e.g. Bliege Bird, Smith, & Bird, 2001; Hawkes, 1991, 1993; Hill & Kaplan, 1988; Sosis, 2000). Hunters in these societies do not have control over the meat they bring to camp, and in some societies there is group-wide sharing (especially at feasts) or at least no significant relationship between what each hunter gives to another household and what he receives from that household (e.g. Bliege Bird, Bird, Smith, & Kushnik; 2002; Bliege Bird & Smith, 2005; Hawkes, O'Connell, & Blurton Jones, 2001a, b; Hill and Kaplan, 1988; Kaplan and Hill, 1985). Thus, meat from big game may be a public good in those societies because it is costly to provide (at least in terms of the opportunity cost of acquiring smaller, non-shareable resources and game), and many benefit from it even if they did not give anything to the hunter. Given the possibility of this and other public goods in ancestral situations, humans may have evolved cognitive mechanisms for dealing with public goods or other collective action problems.

Many laboratory studies have investigated the provision of public goods. Typical experiments use a "public goods game", where participants are given a number of dollars that they can keep for themselves or contribute to a group fund, with the understanding that all contributions get multiplied by some factor (e.g. doubled) before being redistributed evenly among all participants. As long as the multiplier is greater than 1 and less than the number of group members, participants have a selfish incentive to free-ride upon the contributions of others, yet all are worse off if everyone does so (Dawes & Messick, 2000). Participants

usually contribute between 40% and 60% of their endowments in such games, and contributions typically drop with repeated play (Davis & Holt, 1993; Ledyard, 1995). Contributions are especially likely to drop if participants find out that others have contributed less than them, presumably because participants retaliate by also contributing less (e.g. Andreoni, 1995; Fischbacher, Gächter, & Fehr, 2004). Theorists and researchers in evolutionary biology, social psychology, political science, and economics are all interested in the factors that promote cooperation and prevent the drop in contributions.

1.4.2 Selective Incentives for Cooperation: Punishment and Reward

One factor that increases contributions to public goods is the provision of selective incentives, such as punishment for non-cooperation. If participants can punish each other in public goods games by paying money to make others lose money, then they tend to punish low cooperators, and the presence of such sanctions raises cooperation levels (e.g. Caldwell, 1976; Fehr & Gächter, 2000, 2002; Ostrom, Walker & Gardner, 1992; Yamagishi, 1986). In non-laboratory settings, such punishment can include criticism, ostracism, and physical or social threats. Gossip can have "real economic consequences" in stable communities (Fessler, 2002) as it affects one's reputation, and nonmonetary punishment (i.e. social disapproval) raises contributions in public goods games (Masclet, Noussair, Tucker, & Villeval, 2003). In field settings, low contributors tend to inspire more disapproval and receive more criticism than high contributors (Barr, 2001; Cordell & McKean, 1992; Price, 2005), although very high contributors do

sometimes receive punishment (Barr, 2001; Gächter & Herrmann, 2004). Boyd and Richerson (1992) mathematically proved that cooperation can evolve when punishment is possible because defectors are prevented from free-riding on the cooperation of others. Indeed, some form of mutual monitoring and sanctioning is crucial in preventing overexploitation of common resources (Ostrom, 1990). Punishment of free-riders has been dubbed "altruistic punishment" because it is individually costly to perform, yet all group members benefit when free-riders start to cooperate (Fehr & Gächter, 2002).

Altruistic provisioning of public goods can also evolve if contributors are rewarded for their cooperation (Sigmund, Hauert, & Nowak, 2001). Milinski, Semmann and Krambeck (2002a; Semman, Krambeck, & Milinski, 2004) had participants play an experimental game where they alternated between the opportunity to donate money to other players (an indirect reciprocity game from Wedekind & Milinski, 2000) and the opportunity to donate to a public good. They found that people donated more often in the indirect reciprocity game towards people who had contributed to the public good. Clark (2002) and McCusker and Carnevale (1995) found that people were willing to pay into a fund that rewarded the highest public good contributor in their group. Sefton, Shupp and Walker (2002) found that people reward those who contribute more than average to public goods, and van Soest and Vyrastekova (2004) found that people reward those who cooperate by showing restraint in harvesting a common pool resource. Milinski, Semmann and Krambeck (2002b) showed that people who donated money to a

charity were given more money and selected as potential group leaders more often than people who donated less to charity, even when the rewarders did not benefit directly from this. These results all clearly show that people will sometimes voluntarily reward those who help provide public goods.

Status is one potential reward for altruism. Recipients may pay particular attention to altruists such that altruists are prioritized in group member's attention structure (Hawkes, 1993). Fershtman and Weiss (1998) provided a model showing that gaining status is an effective motivator of altruism given that people care about status, and there is good reason why they should. High status people (relative to low status people) are imitated and deferred to more often (Henrich & Gil-White, 2001), receive better offers in bargaining and sharing experiments and in simulated markets (Ball & Eckel, 1996, 1998; Ball, Eckel, Grossman, & Zame, 2001; Commins & Lockwood, 1979), have greater control of resources (Betzig, 1988; Ellis, 1993), and are more likely to survive population crashes (Boone & Kessler, 1999). Furthermore, high status men have more wives and children than low status men (e.g. Mealey, 1985).

In support of the idea that altruists can gain status from their acts, Price (2003) found that Shuar hunter-horticulturalists (of Ecuador) who participate in collective action are likely to be high status group members, although the correlational data do not allow us to infer causation in either direction. Gaining or maintaining status is generally accepted to be the function of some large scale demonstrations of generosity, such as the potlatch tradition among the Kwakiutl

of coastal British Columbia, where much food and many gifts were given away (Goldman, 1937; Rohner & Rohner, 1970; but see Drucker & Heizer, 1967). Hawkes (1990) argued that men will become big-game hunters if those who provide collective food are granted higher status, sexual access, or favourable treatment for their children as rewards. Hawkes presented a mathematical model demonstrating that a male strategy of "showing-off" by providing collective food is evolutionarily stable, and seems to match the behaviour of male Ache and !Kung foragers (reviewed by Hawkes, 1990). Hill and Kaplan (1988) found that good Ache hunters had more extra-marital affairs and more illegitimate children than poor hunters did, and the former's children were more likely to survive to maturity. Hill and Kaplan argued that extra marital affairs and better treatment of hunter's children could serve as rewards given to hunters to motivate them to stay in the group and continue to provide the community with food.

1.4.3 Second-Order Free-Riding

Although punishment and reward sound like solutions to the free-riding problem, several researchers have noted that the provision of selective incentives is a public good itself, because those who provide this "second-order public good" pay a cost that "second-order free-riders" (i.e. non-punishers and non-rewarders) do not (e.g. Hawkes & Bliege Bird, 2002; Oliver, 1980; Ostrom, 1990; Yamagishi, 1986). Rewards involve giving up something (be it time, effort, resources, or relative status) to a cooperator. Punishments such as criticism, ostracism, and physical or social threats all carry risks to the punisher in the form

of potential retaliation, enmity, or the loss of partnership or personal reputation. People who are not motivated to reward or punish would likely benefit more from those incentives being provided than people who have such motivations and act on them, because the former do not pay the cost of providing incentives and yet still benefit from them being provided by others. If this occurred in ancestral environments, then there would have been selection against punitive sentiments and inclinations to reward in those contexts. Punishing and rewarding could also decrease in frequency within an individual's lifetime if people learn (from experience or by observing others) that providing incentives brings fewer relative gains than not providing them. People should notice and care that non-punishers and non-rewarders are better off than punishers and rewarders given that humans care about their payoffs relative to others (e.g. Bolton & Ockenfels, 2000; Roth, 1995), are sensitive to people taking benefits without paying the appropriate costs (Cosmides & Tooby, 1992), and can learn by observation (Tomasello, Kruger & Ratner, 1993). Thus, punishments and rewards should decrease in frequency both within generations (via learning) and over evolutionary time (as punitive and rewarding sentiments are selected against) unless there is some process that supports the provision of incentives for cooperation.

One possible solution to this second-order free-rider problem is to invoke yet another level of cooperation: second-order punishing or second-order rewarding (Henrich & Boyd, 2001). This involves punishing those who do not provide the second-order public good (i.e. those who do not punish or reward) or

rewarding those who do. To sustain this level of cooperation, we would need to invoke even higher levels of cooperation, and so on *ad infinitum*. However, some theorists (Boyd, Gintis, Bowles, & Richerson, 2003; Henrich & Boyd, 2001) have noted that the fitness cost of punishing free-riders (relative to non-punishing) is less than the fitness cost of cooperating (relative to free-riding). This can occur because: i) once punishment is common, it does not need to be provided often to induce cooperation, just often enough to act as an incentive (Boyd et al., 2003; Henrich & Boyd, 2001); ii) the cost to the punisher may be less than the harm inflicted by punishment (e.g. Gintis, 2000), such that the amount of punishment necessary to induce cooperation costs less than cooperation itself would; and iii) if there are multiple punishers, an individual's share of punishing is less than the amount of punishment necessary to induce cooperation. The second and third of these arguments are also likely to apply to the provision of rewards, such that rewarding cooperators for providing public goods costs less than it would cost to provide the public good. However, there is no experimental evidence to date for the existence of second-order punishment, and two recent studies found a conspicuous lack of second-order punishment (Kiyonari & Barclay, 2005; Kiyonari, Shimoma, & Yamagishi, 2004).

1.5 Does Group Selection Solve Second-Order Free-Riding?

If the fitness cost of providing higher-order cooperation is relatively small, then other selection pressures do not have to be very strong to overcome the

fitness disadvantage of providing incentives, such that overall there would be selection for altruism and providing incentives. Henrich and Boyd (2001) note that humans tend to conform to the most common behaviours in their groups, and suggest that the presence of this "conformist transmission" of behaviour would cause norms for punishment to spread within groups if there is little disadvantage to punishment at higher-order levels. Once punishment or other incentives become common within groups that are relatively stable, then they can spread via group selection because groups that provide incentives for cooperation will tend to have higher levels of cooperation than groups that do not, causing the former to have higher fitness than the latter. Indeed, a computer simulation by Boyd and colleagues (2003) showed that altruistic punishment could evolve via group selection even though it is individually costly. There are a few ways in which this could occur. More cooperative groups could outcompete and then replace less cooperative (and hence, less successful) groups (Boyd et al., 2003; Gintis, 2000). More cooperative groups could simply reproduce faster than less cooperative groups, causing an overall increase in the frequency of altruists in the population (Sober & Wilson, 1998; Wilson, 1998, 2004). In a third model, cooperation spreads as less successful groups imitate the cooperative and punitive norms of more successful groups (Boyd & Richerson, 2002).

In order for between-group selection to be stronger than the within-group selection against altruism (including the provision of incentives), Gintis's group selection model (2000; see also Boyd et al., 2003) assumes high rates of group

extinction and relatively low rates of gene flow between groups such that the cooperative groups are not overrun by free-riders. Given that between-group conflict can involve men capturing wives and marriages between members of different villages are not uncommon (Chagnon, 1988), I feel that the second assumption is unlikely to reflect ancestral conditions. Wilson's (1998, 2004) model explicitly relies on high levels of gene flow so that altruism spreads between populations faster than it is selected against within-populations. In some versions of this model, some individuals are willing to produce goods for their groups that end up being shared (such as hunted meat) because it is better to have a share of the good than to have nothing at all, and there is a polymorphism of providers and scroungers that is maintained by the opposing forces of individual and group selection. However, Harpending (1998) has noted that Wilson's model works equally well when individuals dispose of excess resources as when they share those resources, such that group-beneficial acts and group-level advantages are not necessary components of such models, making it unlikely that such a mechanism could select for group-level adaptations. Also, none of these group selection models can account for altruism that occurs in isolated groups with little gene flow to other groups and little opportunity to imitate other groups or compete with them and replace them, as might occur on islands.

Furthermore, it is unclear how cooperation, punishment, and rewarding become common within groups in such group selection models. Conformist-transmission (Henrich & Boyd, 2001) and cultural group selection (Boyd &

Richerson, 2002) both rely on cooperation and punishment being the most common behaviours within groups. Unless all group members simultaneously agree to adopt such norms (possibly after discussion), such behaviours would have to be started by a small number of individuals and then spread despite opposing selection pressures. Although Henrich and Boyd rightfully note that the fitness disadvantage of punishing is not large once punishment is common and everyone cooperates, the cost of punishment is high when punishers are rare, and is especially high when non-cooperation is the norm (Oliver, 1980). Genetic drift would have to be very strong to overcome the selection against altruism and altruistic punishment and make them the most common behaviours. Prestige-based imitation (imitating the most successful group members, Henrich & Gil-White, 2001) alone cannot account for the presence of cooperation and punishment unless altruists and punishers already have high status (Henrich & Boyd, 2001), and this begs the question of why they would tend to have high status (but see section 1.6.3 on costly signaling). Furthermore, learning-based models (conformist-transmission, prestige-based imitation, cultural group selection) do not specify how people know which behaviours to copy, so people would have to copy all of the behaviours that others perform,[1] which is not realistic given how many different types of behaviour people perform in a typical day. Finally, invoking social pressures to maintain the presence of punitive norms

[1] The mind may have mechanisms to prepare it to specifically learn altruism and punishment from others, but that would involve more specificity than the general mechanisms proposed by conformist-transmission and cultural group selection, and would require natural selection specifically for those mechanisms.

raises a problem: if a group starts with non-punishment as the normative behaviour, then those same social pressures would also likely prevent the spread of punishment such that it would never become common within groups. This would not be the case if humans possessed a predisposition to only adopt punitive (and not non-punitive) norms, but invoking such a predisposition to explain the spread of punishment creates a circular argument because it relies on humans having the very predispositions that the argument is trying to explain.

1.6 Individual-Level Benefits for Being Altruistic

If altruists (including incentive-providers) receive individual benefits for their acts, then this could make up for the cost of such behaviours and select for altruistic and punitive sentiments. Once altruism, punishment, and rewards are common in groups, they could indeed spread by group selection, but the group-level benefits would be incidental by-products of mechanisms that were designed to bring individual-level benefits. Although there may be group selection involved in the proliferation of altruism between groups, the possibility of group selection does not necessitate that altruism is a group-level adaptation, as Wilson has claimed (1998; Sober & Wilson, 1998).

1.6.1 Indirect Reciprocity Revisited

Providers of public goods may benefit from indirect reciprocity from other group members. This could stabilize collective action, because the second-order free-rider problem would be solved if people who do not provide rewards are

treated as defectors in a system of indirect reciprocity. Thus, altruism towards one's group (such as the provision of public goods) would be like any another cooperative norm that one must uphold in order to receive the benefits of generalized exchange. Panchanathan and Boyd (2004) provide a mathematical model whereby agents can choose to link collective action to a system of indirect reciprocity. In their model, public good providers have a good reputation when they start interacting in an indirect reciprocity system, and free-riders start with a bad reputation. They show that providing the public good *and* discriminating against collective action free-riders constitute an evolutionarily stable equilibrium. Milinski and colleagues (2002a, b; Semmann et al., 2004) found that people rewarded those who cooperated in public goods games. There is not yet any published work showing that people discriminate against those who do not reward public good providers, but Kiyonari and Barclay (2005) found that rewarders receive more benefits than non-rewarders, which is consistent with Panchanathan and Boyd's model. Also, Price (2003) found that a man's respect for public good providers was correlated with his status among Shuar villagers, which suggests that rewarders are also rewarded.

Indirect reciprocity can explain at least one feature of groups that other researchers might argue supports a group selectionist account of human evolution. Researchers have long known that people show favouritism toward ingroup members. People rate ingroup members more positively and cooperate with them more than with outgroup members (Messick & Brewer, 1983), and they accord

ingroup members more money in monetary-sharing experiments (e.g. Billig & Tajfel, 1973; Tajfel, Billig, Bundy, & Flament, 1971), even if the "groups" are created in a laboratory based on arbitrary and ephemeral characteristics. Such behaviour would obviously benefit one's group, and may sound like a group-level adaptation. However, ingroup favouritism appears to be based on an implied system of indirect reciprocity such that people provide benefits to their ingroup members in the hope or expectation that their ingroup members will also give benefits to them (Yamagishi, 2003; Yamagishi & Kiyonari, 2000). Ingroup favouritism disappears when a person's payoff cannot be affected by others' decisions (Karp, Jin, Yamagishi, & Shinotsuka, 1993). Rabbie, Schot, and Visser (1989) showed that outgroup favouritism occurs when a person's payoff depends upon the decisions of outgroup members instead of ingroup members. When Jin and Shinotsuka (1996, cited by Yamagishi, 2003) controlled for expectations of reciprocity in a Prisoner's Dilemma, there was no ingroup bias. Similarly, expectations of reciprocity overwhelmed and eliminated ingroup effects in Prisoner's Dilemmas with sequential (as opposed to simultaneous) decisions (Yamagishi & Kiyonari, 2000). Thus, it appears that expectations of reciprocity account for ingroup favouritism better than a hypothesized group-level adaptation would.

1.6.2 Assortative Interactions

Assortative interactions provide another potential benefit for cooperation. If cooperators can assort with one another and exclude free-riders (a process that

is likely aided by the evolution of language, Smith 2003), then they will receive the benefits of cooperation without being invaded by free-riders. Hawkes (1991) suggested that good meat-providers might surround themselves with the best rewarders, and these rewarders would benefit by getting a greater share of the meat than they would if they did not pay attention to the hunter's actions and stay close to him. McCabe, Rigdon, and Smith (2004) found that trust and cooperation levels in an experimental trust game rose significantly when cooperators were matched with other cooperators, and dropped when pairings were random, even though participants did not know how they were being matched. Gunnthorsdottir, Houser, McCabe, and Ameden (2000) found similar results using public goods games. Sheldon, Skaggs Sheldon, & Osbaldiston (2000) found that people tended to associate with others who had similar prosocial values, such that when people brought their friends to a laboratory public goods game, the highly prosocial people did not do worse overall than the less prosocial participants because the former tended to be in more cooperative groups.

Positive assortment of cooperators might not even require much cognitive specialization as long as organisms can detect the difference between being cooperated with and being defected on, because they will tend to go to whichever others provide them with the most benefits, who in turn will do the same. Eventually, the best cooperators will end up with each other and the rest have to make do with whomever is left. Although groups of cooperators do better than groups of non-cooperators when there are assortative interactions, this need not be

considered group selection as Sober and Wilson (1998) advocate. In a model of assortative interactions, each individual is doing what is in his/her self-interest by assorting with the best cooperators available, and the incentive is not to cooperate in order to benefit the group, but to stay in the cooperative group. Thus, individual-level selection provides a better account of the origin of assortative interactions, and the requisite adaptations would follow from individual level benefits and costs instead of group-level benefits and costs.

1.6.3 Costly Signaling

Zahavi's (1975, 1977a, b) idea of costly signaling simultaneously explains the existence of extravagant signals (such as some forms of altruism) and provides a mechanism to maintain the honesty of signals despite conflicts of interest between signalers and receivers. When a conflict of interest exists between signalers and receivers, signalers have an incentive to send dishonest information that would cause receivers to behave in a way that is beneficial to the sender. How then can a signaler convince a receiver that the signal is honest, and when can receivers trust the information they receive? If individuals who possess a hidden quality are able to tolerate costs that others cannot, then any organism that does accept such a cost (a "handicap") must possess that quality. The presence of high cost signals ensures the honesty of signals if sending such a signal is impossible or not worth the cost for low quality individuals. Zahavi and Zahavi (1997) gave the example of gazelles "stotting" when faced with predators; instead of running away immediately, some gazelles will pause and make vigorous, energetically

costly leaps into the air. Zahavi and Zahavi argued that this is a signal to the predator implying, "Look how vigorous I am; I can afford to take this time and energy and I can jump this high. Don't bother chasing me because you won't catch me.[2]" The predator attends to the signal in order to avoid an energetically costly but fruitless chase, and the stotting gazelle benefits from also avoiding that long chase. Only a fast gazelle can afford to take the time and energy to jump instead of running. Even if slow gazelles could stot, they would be better off running than stotting because the predator might decide to test the honesty of the signal. Thus, only honest signals are performed, signalers are selected to impose extravagant costs upon themselves to prove the honesty of their signals, and receivers are selected to attend to the costly signals in order to gain important information about the signaler.

1.6.3(1) Altruism as a Costly Signal of Abilities and Resource

Costly signaling theory can be applied to altruism. Altruism, by definition, is a costly act. However, the same altruistic act can be differentially costly for individuals with differing qualities, or differentially beneficial for different individuals, such that it is worth it for those of high quality to perform a given act but not worth it for those of low quality. This can explain extravagant donations to charity or lavish examples of sharing (Boone, 1998), especially competitive forms of sharing such as Kwakiutl potlatches (Goldman, 1937; Rohner & Rohner, 1970; but see Drucker & Heizer, 1967). For example, when billionaires such as

[2] This "translation" of the signal is a paraphrase from Dawkins (1976).

Bill Gates give millions of dollars to charities, they demonstrate not only that they possess millions of dollars, but also they can spare that much. Even if an ordinary person could acquire millions of dollars, it would not be worth it to donate that much to charity because any benefits the person receives from that act would be unlikely to outweigh the debt they would accrue or the opportunity cost of spending that money elsewhere. Although the act has the same absolute cost for billionaires and non-billionaires, it does not impose as much of a "fitness" cost on the billionaires. As another example along the same lines, if I jump into a river to save a baby, I am demonstrating (although probably inadvertently) that I have the physical ability to do so. Others who could not handle the river (let alone while carrying a baby) would be more likely to drown if they tried, so the act is less costly to me than it would be to a weaker swimmer. In both these cases, the altruists can benefit from having others know about their underlying quality (wealth or physical ability in these examples), and observers benefit from knowing the altruists' qualities and choosing to mate with them, cooperate or ally with them, or defer to them.

Sending a costly signal need not be intentional, because observers may infer individual quality from an act that a signaler would perform anyway (Lotem, Wagner, & Balshine-Earn, 1999). For example, I may jump into the river because it is my baby and I have a genetic interest in the child's welfare, but observers can still infer my physical abilities from the act. If I benefit from being observed, then the resulting change in the observer's behaviour towards me could provide

selection pressure for altruism towards less related or even unrelated babies, or for an increased level of altruism towards those babies, in order to demonstrate my abilities. Similarly, signaling benefits can create a selective pressure for altruism not only towards reciprocators, but also towards people who are unlikely or unable to reciprocate (Lotem, Fishmann, & Stone, 2002) or even non-human entities such as organizations to "save the environment".

Gintis, Smith, and Bowles (2001) made a formal model of costly signaling via altruistic acts. They showed that providing benefits for others can function as an honest signal of individual quality provided there is sufficient variation in quality and not too many high quality individuals. As the proportion of high quality individuals increases, this divides the benefits of signaling among more people, such that the expected benefit from signaling decreases. Their model supports previous theoretical work, and shows that signaling by high quality individuals (and not signaling when low) is stable when the expected benefits of signaling (which depend on the proportion of high quality individuals) are greater than the cost to high quality individuals yet less than the cost to low quality individuals. Those who perceive such signals will attend to them and mate or ally with the signalers, but the fact that doing so rewards the signaler is incidental since the perceivers ally or mate with signalers because they are acting in their best interest (Bliege Bird, Smith, & Bird, 2001; Hawkes & Bliege Bird, 2002; Smith, Bliege Bird, & Bird, 2003). Females benefit from mating with men who signal high quality because females seek high quality mates, and anyone can

benefit from allying with those who demonstrate the physical skills or coalitional support necessary to acquire large resources. Men may defer to good hunters because the physical skills demonstrated by hunting may be similar to those used in fighting, and it pays to avoid fights with better competitors (Bliege Bird & Smith, 2005).

Although many different kinds of costly behaviour could be used to signal quality (for a review, see Bliege Bird & Smith, 2005), prosocial signals are especially good because they can also signal a person's willingness to share with others (Gintis et al., 2001; Tessman, 1995). Also, signaling by providing public goods increases the "broadcasting efficiency" of the signal, because receivers will pay attention not only to acquire information about the signaler but also in order to receive a share of the public good (Hawkes & Bliege Bird, 2002; Gintis et al., 2001; Smith & Bliege Bird, 2000). Thus, prosocial signals can attract a larger audience per unit of effort than other costly signals (Boone, 1998; Smith & Bliege Bird *in press*). If individuals are competing with each other to attract the best mates and allies and to deter others, and they are using altruism as a costly signal of quality, then they may compete to be the most altruistic group member (Roberts, 1998). In a more general sense, this could occur whenever reputational benefits are a limited resource such that some group members benefit more from signaling altruism than others who are not as altruistic. Some primate researchers have suggested that baboons compete to groom the highest-ranking group members (Barrett et al., 2000), but I know of no experiments on competitive

altruism in humans conducted or published before the work presented in Chapter 2 of this thesis.

Field researchers have begun to find some evidence or potential examples of public goods provision (mostly hunting) being a form of costly signaling. One modern example is donations to charities or alma maters, in which people demonstrate that they have money to spare. Harbaugh (1998) developed a model whereby people gain prestige for donations to charity, and argued that charities report donations in order to give these prestige-seeking philanthropists a motivation to give more. This motivation can even be exploited by reporting donations in categories (e.g. "$100-$200), such that donors will increase their intended donations in order to get the prestige of being in the next highest category. Harbaugh presented some evidence that donations do tend to increase to match the monetary categories.

Hosting feasts or potlatches can signal the resources of the host (Boone, 1998), and the host's ability to benefit allies (Smith, 2003). Among the Kwakiutl of coastal British Columbia, competitive potlatching increased the standing of both parties in the eyes of observers, and failing to match the size and generosity of other potlatches was considered shameful (Goldman, 1937). When more resources flowed into the Kwakiutl economy due to European influence, the size and frequency of potlatches increased (Drucker & Heizer, 1967), as one would predict if chiefs were trying to outcompete others by signaling their relative wealth. Drucker and Heizer argued that formal positions of status were rarely

gained by throwing potlatches, but potlatching was necessary to confirm or validate such positions of status. Although formal positions of status were rarely gained from potlatching, informal prestige and esteem could clearly be gained from magnanimity in potlatches.

Hunting big-game may function as an honest signal of a hunter's physical abilities. Hunting requires skill, such that there are consistent individual differences between hunters' rates of acquisition, and a man's skill is a better predictor of the amount of meat he catches than the time he spends hunting (Kaplan & Hill, 1985; Hawkes et al., 2001b). In fact, good hunters (those with high acquisition rates) tend to magnify the differences between themselves and poor hunters by spending more time hunting, resulting in even greater differences in meat provisioning. There is evidence that good hunters show off their talents and others attend to the signal to ensure a share of the meat, because good hunters catch more meat when near their village (where there is an audience), and not-as-good hunters are more likely to be present on "bonanza" days when much food is brought in (Dwyer & Minnegal, 1993). Wood and Hill (2000) presented drawings of two different hunting groups (both with single women present) to Ache hunters, and found that men without dependent offspring expressed a preference for associating with the less successful group such that they could be the best hunter in the group. Men with dependent offspring showed the opposite preference. This suggests that the men without dependent offspring wanted to

show off their skill, whereas men with dependent offspring were more concerned with the amount of food that would be available for those offspring.

Torch fishing on the Ifaluk atoll requires much more effort for a smaller return than other forms of fishing, and is a good predictor of a man's productivity at other forms of fishing. Men who torch fish are on average younger and less likely to be married than those who do not, so torch fishing could function as a signal of a man's work ethic to potential mates (Sosis, 2000). However, no data are yet available on whether torch fishers do benefit from these costly displays.

The best-studied potential example of costly signaling in humans occurs among the Meriam of Australia's Torres Strait. Some males hunt turtles to provide for feasts, even though turtle-hunting has a much lower return rate than other types of fishing, is potentially risky, and is costly because of the necessary gasoline for the boats. Furthermore, the hunters do not control the distribution of meat or get any more meat than other people (Smith and Bliege Bird, 2000). Hunting does require the resources to fund a hunt, leadership skills on the part of the hunt leader, and physical skills on the part of other hunters, so turtle-hunting can signal abilities, local knowledge, and resources. Turtle hunting is less likely to occur in turtle nesting season when turtles can be easily collected off the beach, because providing turtle meat in that season is no longer a costly signal of hunting ability and resources (Bliege Bird et al., 2001). During the non-nesting season (when hunting is an honest signal), turtle-hunting teams are composed of better hunters than during the nesting season (when signals can be faked because turtles

can be easily collected). Community members do seem to attend to the signal, because all group members know who the best turtle hunters and spear-fishers are (both are putative costly signals), yet there is no such consistency about who is the best at non-costly shellfish collection or collection of turtles from beaches.

This signaling does appear to benefit the turtle hunters because hunters have higher age-specific reproductive success than non-hunters (Smith et al., 2003). Hunters have more mates, and harder-working mates, than non-hunters. Hunter's wives have higher age-specific reproductive success and are more likely to have at least one child than wives of non-hunters, suggesting that women benefit from mating with hunters. Turtle-collectors fare no better than non-collectors, suggesting that the effect is specific to hunting and is not caused by others reciprocating the provision of meat. Skill at other things like fishing, dance, politics, or wooing women, do not seem to provide higher reproductive success. Hunters have higher reproductive success than their non-hunting brothers, which provides some evidence (albeit not very strong) that the "benefits" of hunting are not epiphenomena of hunters simply having better phenotypes that cause hunting and high reproductive success. Smith (2004) discusses different explanations for the high reproductive success of hunters, and argues that the data best support the hypothesis that hunters benefit from honestly signaling their abilities.

1.6.3(2) Altruism as a Costly Signal of Cooperative Intent

Clearly, not all altruistic acts are sufficiently difficult or costly such that they could be costly signals of abilities or resources. Some altruistic acts are easy

or cheap enough that almost anyone could perform them if they had the desire to do so. However, generosity could signal cooperative intent or commitment to a common project, such that altruism is not worth the cost for those who intend to defect on cooperative partners (Smith, 2003). Smith and Bliege Bird (*in press*) note that a signal of cooperative intent can be worth it for someone who intends to make up those costs over time by cooperating in prolonged interactions. Observers should seek these cues of cooperative intent in order to avoid being cheated in social exchanges, especially when there is a reasonable chance of encountering a non-cooperator in the population (McNamara & Houston, 2002). Some might argue that such a signal need only be sent at the start of a relationship. However, Bliege Bird and Smith (2005) suggest that repeated signaling may be necessary if a person's past condition (or willingness to cooperate) is not fully predictive of future condition (or willingness to cooperate), if cessation of signaling could be interpreted as a cessation of willingness to cooperate in the future, or if there is noise in the system such that the presence or strength of a single signal is not always easily determined and multiple signals are required to accurately judge cooperativeness.

Few studies have tested whether altruism signals a willingness to cooperate. Although they were not directly testing that hypothesis, Kurzban and Houser (2005) found that people who cooperated in a public goods game with one group were likely to cooperate with other groups, such that people could be consistently categorized as cooperators, free-riders, and conditional cooperators.

This is a necessary condition for signaling cooperative intent because it shows that people who cooperate at one point in time are more likely to cooperative at a later point. Clark (2002) and Sefton et al. (2002) found that people who contribute to public goods tend to be the ones who reward others for contributing, which also suggests that contributions in public goods are predictive of future cooperative behaviour.

Wedekind and Braithwaite (2002) had participants play an indirect reciprocity game and then a Prisoner's Dilemma in dyads, and found that people were more likely to cooperate in the Prisoner's Dilemma with those who had been generous in the indirect reciprocity game than with those who had been less generous. Numerous studies have shown that people are much more likely to cooperate in social dilemmas when they believe that others will also do so (e.g. Komorita & Parks, 1995; Liberman, Samuels, & Ross, 2004; Messick & Brewer, 1983; Smeesters et al., 2003). However, it is unclear whether participants in Wedekind & Braithwaite's study cooperated with generous people because they wanted to reward the generous people, or because they believed that the less generous people were less likely to cooperate in the Prisoner's Dilemma and they did not want to be "suckered" by cooperating when their partners defected. Thus, it is unclear whether this study specifically supports the notion that people will interpret altruistic behaviour as signal of future cooperative intent, but the data are consistent with this idea.

Albert, Güth, Kirchler, and Maciejovsky (2002) showed that people who gave large amounts of money to a charity were trusted more in trust games and cooperated with them more often in Prisoner's Dilemma games than people who were less generous. Furthermore, all other players preferentially trusted them except the people who had donated the least amounts to charity. Albert et al.'s (2002) results also suggested that highly generous people were more discriminating about whose trust they repaid. When paired with other generous people, they cooperated more often than moderately generous or relatively stingy people, but when paired with stingy people, they cooperated less than moderately generous people did. Thus, the altruists were not more trustworthy overall, but were more trustworthy towards generous people, so the altruism could be an honest signal of cooperative intent towards other generous people. These results do not show that people actively signaled their altruism, but did show that people responded to it as if it were a signal.

In an experiment by Keser (2003), people played a series of trust games (from Berg, Dickhaut, & McCabe, 1995) in which one player could send money to a partner, and that money got tripled before the second player decided how much (if any) to return. The first players then gave the second players a positive, negative, or neutral rating. Participants played 20 rounds like this. Players were randomly repaired every round, and had access to their partners' previous ratings. Participants entrusted more money to others when they had access to their partners' reputations than when they did not. Furthermore, participants entrusted

more money to others when they had access to their partners' long-term reputations (i.e. information on average ratings) than when they only knew their partner's short-term reputations (i.e. information on the rating in the previous round only). Keser (2002) also found that participants returned more money to the senders when they could acquire a reputation for doing so, and people tended to trust those who had been trustworthy in the past. These results show that players were concerned about their reputations for trustworthiness, others responded to those reputations, and participants may have behaved in a trustworthy manner in order to gain from partner's trusting behaviour in future rounds. However, this study only used one type of experimental game, and as such it did not show that people would behave cooperatively in one context in order to signal trustworthiness in another context. In Chapter 2, I will present evidence that people do try to use displays of altruism in one context to signal trustworthiness in a different context, and further evidence that participants are more likely to trust people who make high contributions to public goods than others who make lower contributions. Furthermore, I will present evidence that incentives to compete for the best reputation can maintain contributions to public goods better than opportunities for reputation without such incentives.

1.6.3(3) Altruistic Punishment as a Costly Signal

Gintis et al. (2001) suggested that punishment can be a costly signal of individual quality or status, given that dominant individuals are better able to punish subordinate individuals than vice versa (Clutton-Brock & Parker, 1995).

Punishment can invite retaliation, and dominants are better able to withstand such retaliation than subordinates. Also, punishment of a high status individual by a low status individual is likely to be ineffective if the low status person lacks the strength or social power to harm the free rider without doing much more harm to himself. Thus, the honesty of punitive signals is maintained since punishment is less costly and more beneficial for high status individuals because their punishment is less likely to invite retaliation and more likely to be effective. In Gintis et al.'s model, one evolutionary equilibrium is for high quality individuals to punish and low quality individuals to abstain from punishing. There is some field evidence that high status individuals are more likely to criticize free-riders than are low status individuals (Barr & Kinsey, 2002; Wiessner, 2003). Chapter 3 of this thesis investigated the possibility that artificially granting people high status makes them more likely to provide altruistic punishment of free-riders.

McElreath (2003) modeled the effects of reputation in conflict situations. He found that individuals should be more willing to fight over resources when there is a possibility of acquiring a reputation for willingness to fight. Having a tough reputation deters others from escalating conflicts over resources, such that individuals with hawkish reputations are more likely to gain resources without conflict than individuals with dovish reputations. This is allegedly occurring in "cultures of honour" in places such as the southern United States (Cohen, Nisbett, Bowdle, & Schwarz, 1996) where people are very willing to fight to defend their honour. In such places, a tough reputation may be the most effective deterrent

against transgressions because external punishment (e.g. law enforcement) has historically been or still is inadequate.

Similarly, when people sanction free-riders in a group, they may be signaling an unwillingness to be cheated. Observers would then be less likely to defect on anyone who has demonstrated a willingness to punish, whereas they might defect on someone who has conspicuously abstained from punishing, and this can select for punitive sentiments (Brandt, Hauert, & Sigmund, 2003; Hauert, Haiden, & Sigmund, 2004; Sigmund, Hauert, & Nowak, 2001). Similarly, if a person develops a reputation for always being willing to reject unfair offers despite the cost of doing so, then he/she will tend to receive fair offers and will consequently do better than those who are known to accept unfair offers (Nowak, Page, & Sigmund, 2000). Experimental evidence suggests that people are more likely to reject unfair offers when they can acquire a reputation for doing so (Fehr & Fischbacher, 2003) and are more likely to punish people for defecting on third parties when their behaviour could become known to the experimenter (Kurzban et al., 2004). Thus, it may be beneficial to publicly punish defectors in order to deter defections against oneself, and we might predict that people who are most concerned with deterring transgression against oneself.

Finally, altruistic punishment may also signal a person's cooperative intent. When punishing a free-rider is good for a group, it could signal the punisher's trustworthiness, commitment to that group, or concern with fairness. People who demonstrate a concern for fairness in group settings may be more

likely to treat others favourably in dyadic partnerships. Other people would then be more willing to enter a cooperative relationship, and invest more in relationships with people who have demonstrated that they will not tolerate inequity. This would then enable punishers to receive more benefits from cooperative partnerships than non-punishers. Thus, punishment could function as a signal of cooperative intent in the same way that other forms of altruism might, provided that the cost of the punishment is greater than the benefits of cheating someone (which ensures the honesty of punitive signals) and less than the benefits of ongoing cooperation in a partnership (which makes the signal worthwhile to the punisher). Chapter 4 presents the first empirical evidence to test the hypothesis that people will trust altruistic punishers.

If punishment is a signal of quality, status, cooperative intent, or unwillingness to be cheated, then others will attend to that signal because it is in their best interest to do so. Thus, if sanctioning free-riders is a signal of some sort, responding favourably to punishers is immune to the second-order free-riding problem. It is in an observer's best interest to enter cooperative relationships with punishers in order to gain a trustworthy partner. Likewise, observers should avoid cheating such people in dyadic relationships in order to avoid sanctions. If punishers do receive some type of reputational benefit, then trust and respect (or fear) are good candidates. If such reputational benefits translated into tangible benefits in ancestral environments, then this could explain the existence of punitive sentiments.

1.7 Where Does This Leave Us?

Some investigators have argued that humans do not seem to receive individual benefits for many altruistic acts, so altruistic sentiments must have been designed by group selection (e.g. Sober & Wilson, 1998). I believe that this is a premature conclusion, and have described some of the ways in which altruists might benefit from their actions. If there are individual-level benefits for generosity, then this reduces (and possibly eliminates) the necessity of relying on group selection to explain the evolution of cooperative and punitive sentiments. Group-selectionist and individual-selectionist theories make different predictions about how humans behave, and a premature rejection of the latter will prevent us from making many interesting predictions. In this thesis I will test some of the predictions that costly signaling theory might make about human altruism, and present evidence of individual-level benefits that cooperators and altruistic punishers receive.

Chapter 2: Trustworthiness and Competitive Altruism Can Also Solve the "Tragedy of the Commons"

Abstract

The benefits of a good reputation can help explain why some individuals are willing to be altruistic in situations where they will not receive direct benefits. Recent experiments on indirect reciprocity have shown that when people stand to benefit from having a good reputation, they are more altruistic towards groups and charities. However, it is unknown whether indirect reciprocity is the only thing that can cause such an effect. Individuals may be altruistic because it will make them more trustworthy. In this study, I show that participants in a cooperative group game contribute more to their group when they expect to play a dyadic trust game afterwards, and that participants do tend to trust altruistic individuals more than non-altruistic individuals. I also included a condition where participants had to choose only one person to trust (instead of being able to trust all players) in the dyadic trust game that followed the cooperative group game, and contributions towards the group were maintained best in this condition. This provides some evidence that competition for scarce reputational benefits can help maintain cooperative behaviour because of competitive altruism.

Note: This chapter is reproduced under licence from Elsevier Publications from Barclay (2004), which was published in *Evolution and Human Behavior* in volume 25 on pages 209-220.

2.1. Introduction

Altruism towards unrelated individuals has puzzled evolutionary biologists for decades, and several theories provide possible explanations for its existence. Theories of direct reciprocity (Trivers, 1971) and indirect reciprocity (Alexander, 1987) suggest that organisms can succeed by reciprocating altruistic acts towards other altruists. Direct reciprocity occurs when individuals reciprocate generous acts towards others who have been generous to them in the past. Indirect reciprocity occurs when individuals provide benefits for others who have been generous to anyone, and in turn are rewarded for this benevolence by individuals other than the recipients. Many computer simulations and experimental games have shown that some forms of direct and indirect reciprocity can allow for the evolution of altruism, and people actually do engage in direct and indirect reciprocity (see especially Axelrod, 1984; Nowak & Sigmund, 1998; Wedekind & Milinski, 2000; but see also Leimar & Hammerstein, 2001).

However, these theories by themselves cannot account for altruistic acts that cannot be directed towards particular individuals, such as the provision of public goods. A public good is something that people have to incur costs to provide and yet all members of the group benefit from it whether or not they helped provide it (Davis & Holt, 1993), so the public good is open to exploitation by free-riders. Examples of public goods include group protection, irrigation, and any collective action project. Individuals have an incentive to not provide public goods because the benefits of providing them are spread among many people,

whereas only the altruists bear the cost. Thus, the provision of public goods is very much like the classic "tragedy of the commons" situation introduced by Hardin (1968). One would expect that altruism in such situations would be selected against and yet many studies demonstrate that humans are willing to contribute to public goods (e.g. Fehr & Gächter, 2000).

People may be altruistic in these situations if there is a chance that they will earn a good reputation that will later be repaid in direct or indirect reciprocity (Alexander, 1987). Supporting this, Milinski, Semmann & Krambeck (2002a) had participants play an experimental game where they alternated between the opportunity to donate money to other players (an indirect reciprocity game from Wedekind & Milinski, 2000) or the opportunity to donate to a public good. They found that people were more likely to contribute to public goods when they expected future indirect reciprocity games, and that participants donated more often in the indirect reciprocity game towards people who contributed to the public good. However, the rewarding of altruists (one component of indirect reciprocity) is not the only way in which an individual might benefit from a reputation for altruism. People often engage in dyadic relations in which they have to trust another person, and competition to form these cooperative partnerships could also account for the importance of reputation. The present study examines whether humans are more willing to trust altruistic individuals than non-altruistic individuals in a situation where they might be cheated. Alternately, people might not do so, because it would then become possible for an

individual to send a dishonest signal by being generous in order to deceive others into trusting him/her. This would reduce the effectiveness of altruism as a signal of trustworthiness, such that people do not trust altruists any more than non-altruists.

Given that coalitions and reciprocal altruism are integral parts of human interaction and carry great potential benefits and costs, we can expect careful choice of cooperative partners (Cosmides & Tooby, 1992). If altruism can signal a willingness to cooperate in partnerships, then altruistic individuals will be desirable partners (Alexander, 1987; Brown & Moore, 2000). However, no one can interact with all people all of the time, and people tend to form friendships or interact more frequently with some individuals than others. This should create a subtle competition to be more altruistic than others in order to be preferred as an exchange partner (Roberts, 1998). Whether this occurs in humans is an open question, but it may be occurring in non-human grooming partnerships when good reciprocators prefer to interact with each other (Barrett, Henzi, Weingrill, Lycett & Hill, 2000). It is similar to Seyfarth's (1977) model of primate grooming where there is competition to associate with the highest-ranking individuals (see Schino, 2001, for a review). More generally, competitive altruism could occur whenever the most altruistic individual in a group can stand to receive more benefits than other altruistic individuals, whether those benefits be better partnerships or not.

No studies have explicitly tested for the existence of competitive altruism in humans, so the present study investigated whether humans will compete to be

the most altruistic member of a group. I used a Public Goods Game (PGG) where each participant had an incentive to be selfish, but all participants could have done well if everyone was altruistic. Past studies using PGGs have shown that contributions tend to fall by the last round (Davis & Holt, 1993), and especially in the last round if participants know when the last round is. Therefore, the present study tested whether having an incentive to compete for the most altruistic reputation would maintain contributions better than if there were no extra incentive for being the most altruistic individual.

2.2 Methods

2.2.1 Participants

One hundred twenty participants (43 males, 77 females) were recruited using posters around McMaster University campus. The average age of participants was 23.9 (± SD 5.6) years.

2.2.2. General Procedure

In groups of four, participants played a Public Goods Game (PGG) followed by a Trust Game, and each player was given a pseudonym so that he/she could acquire a reputation in the game but still be anonymous. Participants were seated at a table with dividers that prevented them from seeing each other while they made their decisions, and prevented them from seeing each other's decisions. During the experiment, they earned "lab dollars" which would be exchanged at the end of the experiment to Canadian dollars at a rate of 15:1, with a 1 in 36

chance to change them on par for Canadian dollars (to provide further incentive to treat the "lab dollars" like real dollars). Participants were paid individually at the end of the experiment to reduce the chances of their interacting after the experiment.

2.2.3. Public Goods Game

In each of the five PGG rounds, each player was given 10 lab dollars and had the option of contributing any number of these to the public good. Each round, players' contributions were collected using envelopes that had the pseudonyms inside, and these contributions were written on a blackboard beside each participant's pseudonym. The total contributions in each round were multiplied by 1.6, and this new total was divided evenly amongst the participants. After they received their shares, the next round began. At the start of every round, players were told how many rounds of the PGG remained.

2.2.4. Trust Game

After five rounds of the PGG, participants played a version of Berg, Dickhaut and McCabe's (1995) Trust Game. They used the same pseudonym for the Trust Game as they had for the PGG. They were given a total of 30 lab dollars which they could send to other participants. Any money that a player sent was tripled (by the experimenter) before the other player received it, and the recipient could then return any amount to the sender. Players put the amount they wanted to send into an envelope and wrote down the pseudonym of the player that they wanted to send it to. In order to maximize information about recipients' decisions,

each participant indicated how much money he/she would return for each possible amount they could receive regardless of who sent it: Thus returns were not contingent on the identity of the sender. This decision was binding in that the experimenter used this information to calculate the amount of money returned to the senders.

2.2.5. Experimental Conditions

There were three experimental conditions with 10 groups in each condition. (1) In the No-Reputation condition, participants were told they would play another monetary game after the PGG, but they did not know the details of the Trust Game before playing the PGG, and thus did not have any strategic reason to acquire a good reputation. (2) In the Regular-Reputation condition, participants were informed (before playing the PGG) about the Trust Game they would play with each of the other three players. (3) In the Competitive-Reputation condition, participants were informed (before playing the PGG) about the Trust Game they would play with only one partner of their own choosing. Thus, the No-Reputation condition acts as a control, where participants did not know they could gain a good reputation, while the Regular-Reputation and Competitive-Reputation conditions examine the effects of reputation and competition to have the best reputation, respectively.

In the No-Reputation and Regular-Reputation conditions, participants played the Trust Game with each of the other three players, to whom they could send up to 10 lab dollars each. In the Competitive Reputation condition, each

participant played one Trust Game with one other player *of his/her choice*, and could send up to 30 lab dollars to that one player. Thus, participants had the same number of lab dollars to trust to other players in all conditions, but a single "trustworthy" individual could receive more in the Competitive Reputation condition. This creates an incentive to be the most altruistic individual in the PGG in order to be the one trusted with money in the Trust Game.

Although I predicted that groups with high PGG contributions would tend to have higher trust levels in the Trust Game, I did not predict any additional effect of experimental condition on trust levels. PGG contributions were visible on the blackboard throughout the Trust Game in all three conditions.

2.2.6. Practice Rounds

Practice rounds were conducted to familiarize the players with the nature and the procedures of the games, as well as to make the presence of the Trust Game salient in the two reputation conditions. In the practice rounds, participants were instructed on how much to contribute in the PGG and entrust to others in the Trust Game. The amounts were chosen by the experimenter so as not to bias the participants' decisions for or against the experimental hypotheses.

2.2.7. Statistical Analysis

Each group of four players was treated as one unit of analysis because each participant's behaviour affects the behaviour of others in later rounds. Total group contributions in the PGG were analyzed with a Repeated Measures General Linear Model. Any violations of sphericity assumptions were corrected for using

the conservative Greenhouse-Geisser correction. In PGGs and reciprocity games, contributions often drop in the last round if participants know which round is the last, so I also examined the change in contributions between the fourth and fifth rounds.

In the Trust Game, the three conditions were analyzed separately. In the No Reputation condition, the instructions and practice rounds for the Trust Game were given in between the PGG and the Trust Game, causing a long separation between the two and likely reducing any effects of the former on the latter. Both games were played back to back in the Regular Reputation and Competitive Reputation conditions, but the number of people to whom money could be entrusted was different in those conditions. In addition to correlating players' contributions in the PGG with amount they received in the Trust Game, I ranked the players within each 4-person session on their total contributions in the PGG, and ran Within-Group General Linear Model on the amount sent and received by each player in the Trust Game.

2.3 Results

2.3.1. Effects of Reputation in the Public Goods Game

Figure 2.1 presents the results of the PGG. There was no overall significant difference between the Regular and Competitive Reputation conditions ($F < 1$), so these conditions were pooled to compare with the No Reputation condition to test for the effects of reputation on PGG contributions. The two

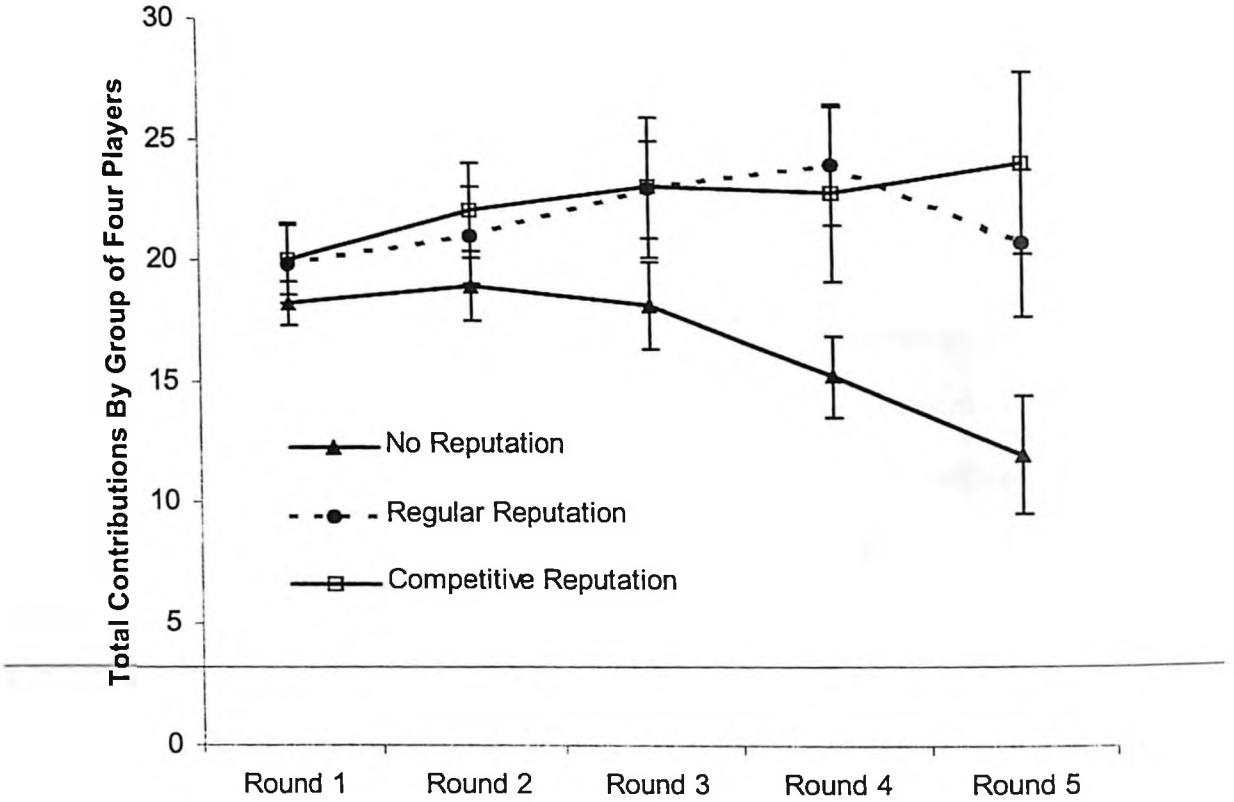


Figure 2.1. Group contributions to the public good dropped in the No Reputation condition (—▲—), but rose in the Regular Reputation (--●--) and Competitive Reputation (—□—) conditions, showing that having an opportunity for reputation makes people more likely to contribute to public goods. Contributions were less likely to drop in the final round of the Competitive Reputation condition than in the Regular Reputation condition, suggesting that when individuals have to compete for the most altruistic reputation, they are more likely to continue being altruistic to the end. The error bars represent standard errors of the means.

pooled conditions with reputation had significantly higher contributions than the No Reputation condition ($F_{1,28} = 5.91, p = 0.022$). There was a significant interaction of reputation with round number ($F_{1.61,45.06} = 3.69, p = 0.042$). This interaction was caused by different linear trends in conditions with and without

reputation (linear contrast analysis $F_{1,28} = 4.69, p = 0.039$), such that contributions tended to drop without reputation and slightly increase with reputation.

2.3.2. Competition for Reputation in the Public Goods Game

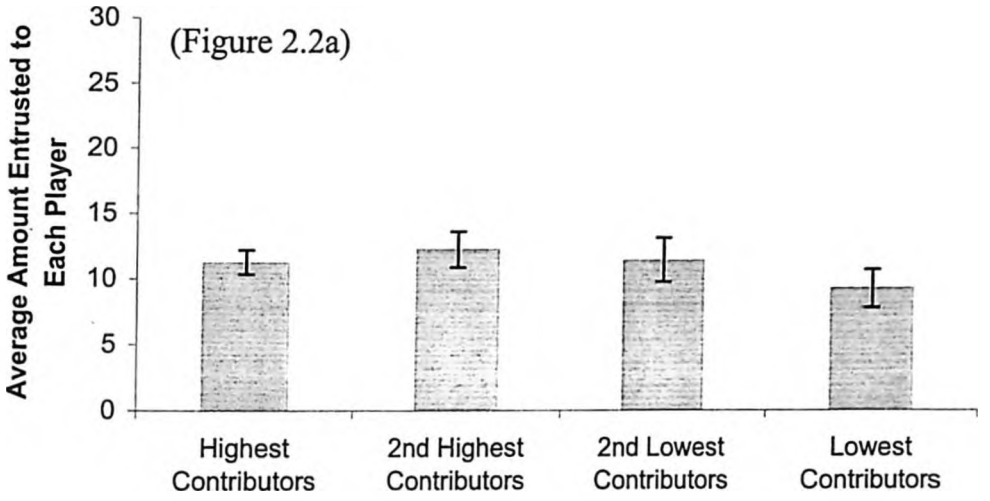
The Regular Reputation and Competitive Reputation conditions were very similar until the last round. In the final round, contributions dropped an average of $3.20 (± SEM 1.40) per group in the Regular Reputation condition, but rose an average of $1.28 (± SEM 1.40) per group in the Competitive Reputation condition (Figure 1). This difference was significant ($F_{1,18} = 5.16, p = 0.036$), indicating that contributions were less likely to drop in the Competitive Reputation condition. Furthermore, this difference was significant even if round 4 is included as a covariate to control for differences in round 4 contributions ($F_{1,17} = 4.85, p = 0.042$).

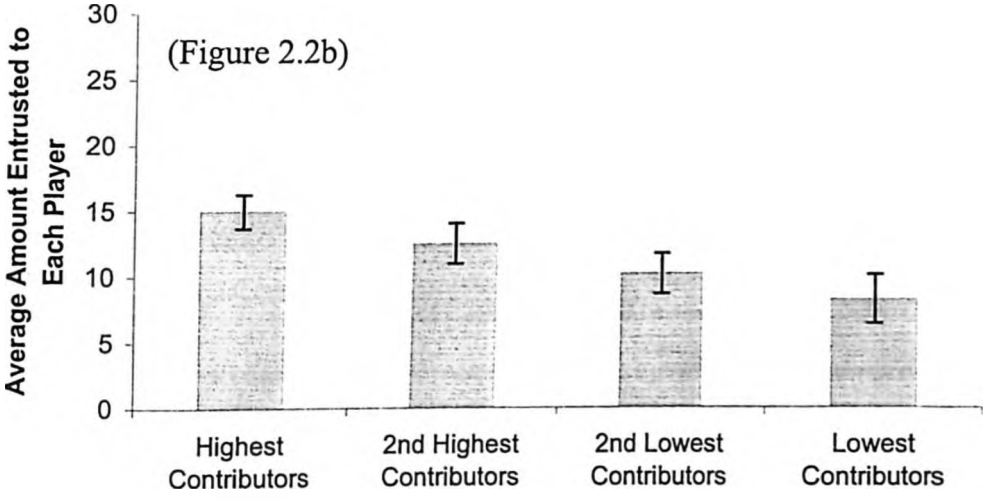
2.3.3. Trust Game

Total amounts sent in the Trust Game were not significantly different across the No Reputation ($M = \$11.6$), Regular Reputation ($M = \$11.9$), and Competitive Reputation ($M = \$10.9$) conditions ($F < 1$). Participants tended to entrust more money to players who contributed more in the PGG in both the Regular Reputation condition ($r_{38} = 0.72, p < 0.001$) and the Competitive Reputation condition ($r_{38} = 0.53, p < 0.001$). In these two conditions, there was a

strong correlation between total contributions in the PGG by all four players in each group and the total amount sent in the Trust Game by all four players (r_{18} = 0.75, $p < 0.001$). Even after factoring out this effect, a player's total contributions still predicted the amount of money he/she was entrusted with in the Trust Game in the Regular Reputation condition (partial $r_{37} = 0.59$, $p < 0.001$) and in the Competitive Reputation condition (partial $r_{37} = .40$, $p = 0.012$). All of the above correlations were positive but not significant in the No Reputation condition (all ps > 0.10).

There were significant differences in how much differently-ranked players received in both the Regular Reputation conditions ($F_{3,27} = 7.60$, $p = 0.001$) and the Competitive Reputation condition ($F_{3,27} = 5.79$, $p < 0.01$; Figure 2.2). In the Regular Reputation condition, the top-ranking PGG contributor in each group of four received significantly more than the bottom-ranking and second lowest contributors (Fs$_{1,9}$ = 47.38 and 9.89, $p < 0.001$ and $p = 0.012$, respectively), and the second highest contributor received significantly more than the bottom-ranking contributor ($F_{1,9} = 5.37$, $p = 0.046$), but there were no other significant differences. In the Competitive Reputation condition, the bottom-ranking PGG contributor in each group of four received significantly less than the other three players (all Fs $_{1,9}$ > 10, all ps < 0.01), but there was no effect of the other rank positions on amount entrusted (all Fs < 1). There were no significant differences in the No Reputation condition, although the effect approached significance ($F_{3,27}$


(Figure 2.2a)
Average Amount Entrusted to Each Player
30
25
20
15
10
5
0
Highest Contributors
2nd Highest Contributors
2nd Lowest Contributors
Lowest Contributors


(Figure 2.2b)
Average Amount Entrusted to Each Player
30
25
20
15
10
5
0
Highest Contributors
2nd Highest Contributors
2nd Lowest Contributors
Lowest Contributors

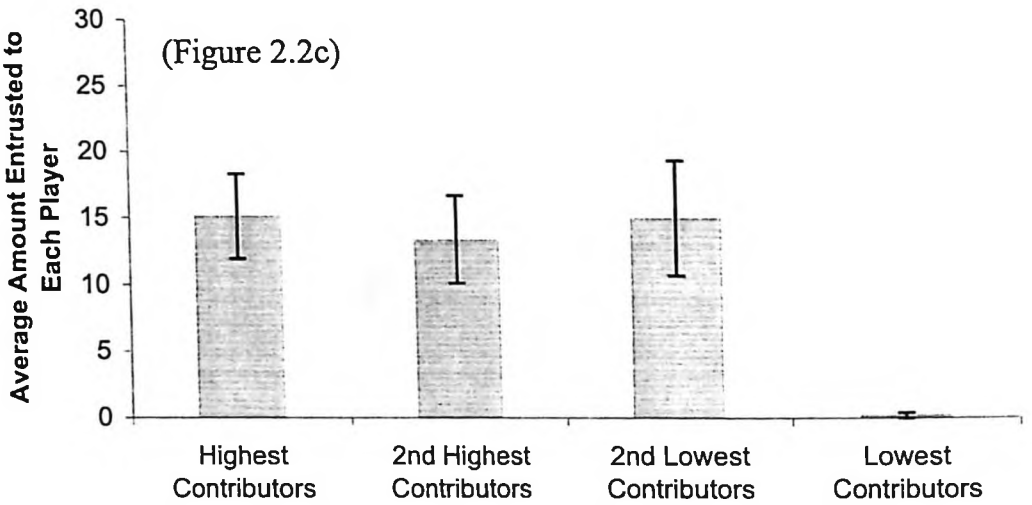


Figure 2.2. Average amount (and standard errors of the means) sent to each player in the Trust Game as a function of their rank as contributors in the PGG. (a) There were no significant differences in the No Reputation condition. (b) In the Regular Reputation condition, the highest-ranking PGG contributors were entrusted with more money than the second-lowest or lowest-ranking contributors, and the second-highest contributors were entrusted with more than the lowest-ranking contributors. (c) In the Competitive Reputation condition, the lowest-ranking PGG contributor received less than the other three players did.

= 2.62, $p = 0.07$) because the lowest contributor received less on average than any of the others ($ps < 0.10$).

These effects were not due to high contributors trusting less (i.e. sending less money) than low contributors in the reputation conditions. After factoring out the total PGG contributions by all four players in each group (as above), there was no significant correlation between an individual's contribution in the PGG and how much he/she sent to others in the Trust Game (partial $r_{77} = -0.01$, $p > 0.92$). Similarly, there were no significant differences among players in a group in how

much they sent in the Trust Game, based on the rank of their total contributions within their group ($F < 1$). The only significant predictor of the proportion of money an individual was willing to return in the Trust Game was the amount that the individual him/herself sent to other players ($r_{78} = 0.52, p < .001$).

4. Discussion

This study shows that people are more willing to contribute to a public good when they can benefit from having a reputation for being altruistic, because contributions were lower in the No Reputation condition than in the two reputation conditions (Regular Reputation and Competitive Reputation). This study also shows that people may be most altruistic when they could benefit from being the most altruistic individual in a group: PGG contributions were less likely to drop in the last round if there were potential benefits to being the most altruistic. Thirdly, the study found that people were more likely to trust individuals who had been altruistic in the public goods game.

This is the first study to show that people will be most altruistic when they might have to compete to be the most altruistic member of the group, and is thus the first experimental evidence for the existence of competitive altruism in humans. A study by Clark (2002) gave participants the opportunity to give money to the most altruistic group member, but this opportunity did little to increase PGG contributions. Clark's participants had no vested interest in giving money, whereas participants in the present study could benefit by trusting if that trust was

repaid. This incentive to trust (and expectation of trust) makes competitive altruism more likely to arise. This could occur in situations where only a subset of all people will benefit from signaling altruism, or when some people will benefit more than others will. Such situations may include (but are not limited to) times when individuals need to cooperate with others, yet cannot or will not form partnerships with all other group members. The most altruistic individuals may attract the best (or the most) cooperative partners (or mating partners if pair-bonding is a type of cooperative relationship where free-riding on a partner's efforts is possible.)

The present study replicates and extends Milinski et al.'s (2002) findings that more people contribute in PGGs when they expect future indirect reciprocity games, and people donate money more often to persons who contributed in the PGG. The present experiment differs from the Milinski et al. study because it shows that reputation effects extend to trust. It does so by pairing a PGG with an experimental game that measures trust rather than indirect reciprocity, and shows another reason why good reputations may be valuable. The present results in the Trust Game are not likely to be the result of rewarding high PGG contributors for a few reasons. The amounts entrusted to other players were higher than one might expect anyone to send as a reward for contributing in the PGG because the average amount entrusted was over one third of the maximum possible, which was more than the endowment in a round of the PGG. In other experiments, people have given about one tenth (Clark, 2002) or one fifth (Sefton, Shupp &

Walker, 2002) of the total amount possible to the highest PGG contributors, whereas in the present study the highest contributors were entrusted with about one half of the amount possible in the two reputation conditions (see Fig 3b,c). Even the amount entrusted to bottom-ranking contributors was greater than zero, and many public goods experiments have shown that people tend to punish low contributors by lowering their payoff (ex. Fehr & Gächter, 2000, 2002; Sefton et al. 2002), not reward them. Finally, one would expect that high contributors in the PGG would be more likely to be the people who give to others in indirect reciprocity, as found by Clark (2002). This was not the case in this experiment because there was no relationship between the amount that an individual contributed and the amount he/she entrusted to others (after group contributions were factored out).

There was, however, a positive relationship between the PGG contributions of each group and how much its members sent in the Trust Game, indicating that cooperation may have facilitated partnership formation. Despite this, amounts sent in the Trust Game were just as high in the No Reputation condition as in the two reputation conditions, despite the lower PGG contributions. This may be because there was a reason to discount the honesty of an altruistic signal in the Reputation conditions: some participants could have contributed in the PGG in order to deceive others into trusting them. This is especially true in the Competitive Reputation condition, where the incentive to make such deceptive contributions was highest because the potential payoffs of

being trusted were highest. Also, the potential cost to the truster was greatest in this condition because he/she could not spread out the risk by trusting more than one person. However, this discounting of altruism should not eliminate the incentive to try to gain the best reputation, because people might not want to risk being passed over as a partner in favour of more altruistic individuals.

If people are sensitive to the possibility of dishonest signals of altruism, then they should vary their trust according to the costs of being cheated and the potential benefits to a deceptive signaler. Since both of these were greatest in the Competitive Reputation condition, a signal of altruism would have been least effective in this condition. This might explain the surprising result that participants did not send more money to the highest contributors than to the second and third-highest contributors in the Competitive Reputation condition (Figure 2c). It also may explain why there was a slightly stronger correlation between a player's contributions and what he/she was entrusted with in the Regular Reputation condition than in the Competitive Reputation condition. One might predict less wariness outside of the laboratory because pairing with more than one person reduces the cost of being cheated by any one individual. Furthermore, deceptive signaling may be less common outside the laboratory because the cost of altruism could be sufficiently high, making it worthwhile only for individuals who will not defect immediately in a cooperative relationship. This was not the case in the present study because the Trust Game was played only once. Despite the possibility of dishonest signals in this experiment, participants

may have felt that predicting others' trustworthiness from their PGG contributions was better than using no information about other players' behaviour.

There are some limitations of the present design that may have hindered the search for competitive altruism. Anything that reduces the likelihood of money being entrusted to other players also reduces the incentive to compete for the best reputation because the benefits of a reputation depend on being trusted with money. There are a few things that might have reduced the likelihood of money being sent in the Trust Game. First, there was no guarantee that players would entrust any money to anyone, and each player should have recognized that and realized that he/she might not receive any money from other players. Secondly, participants had to indicate how much money they would return if they were entrusted with each possible amount, and they could not make these returns contingent upon the identity of the sender. This was done to ensure that the effects of having a good reputation did not cause a ceiling effect in contributions by having high contributors benefit from having money preferentially sent and returned to them. However, this design meant that if players wanted to avoid returning money to those who had contributed little in the PGG, they could only do so by returning relatively little to anyone. Players may have feared that the others would return relatively little, and thus not entrusted as much money as they would have if they could respond differently to each person. Thirdly, and most importantly, there may have been little incentive in the Competitive Reputation condition to be the second- or third-most desirable trust partner, because

participants could only trust one person. Thus, low contributors may have felt that there was no benefit to having a reputation and reduced their contributions in the PGG, affecting the entire group. In real life, the second- and third-most trustworthy people in a group might also benefit from good reputations because people can form multiple partnerships. Despite these limitations, the present study provides evidence for competitive altruism, which suggests that this is a potentially fruitful area for future research.

Chapter 3: Effects of Social Status on Cooperation and Punishment in Public Goods Games

Abstract

If generosity and altruistic punishment sometimes function as costly signals of social status, then perceptions of high status should induce people to be more cooperative and punitive. I used false feedback on a quiz to attempt to manipulate participants' perceptions of their status relative to other group members. After receiving the false feedback, participants then played a public goods game (either with each other or against pre-programmed computer "players") with the option to punish one another. Assigned quiz ranking did not affect levels of generosity or punishment, nor did they affect perceptions of self-esteem or social rank as measured by post-experiment questionnaires, which suggests that the status manipulation was ineffective. Among participants who interacted with the computer "players", higher self-esteem scores were associated with higher totals spent on punishment. Finally, physical proximity to the experimenter seemed to increase public goods contributions and punishment despite the anonymity of participants' decisions.

3.1 Introduction

Humans regularly cooperate in large groups, and are a very cooperative species compared with other animals. Group cooperation is often a social dilemma; everyone in the group is better off when others do so, but the personal benefits of cooperation do not outweigh the personal costs so there is an individual incentive to defect on others (Dawes & Messick, 2000; Messick & Brewer, 1983). Social psychologists, economists, anthropologists, and evolutionary biologists all study cooperation in social dilemmas, and the provision of public goods has recently had much investigation. Public goods are something that individuals have to expend time, effort, or money to provide, and all group members (even the non-providers) benefit from their provision (Davis & Holt, 1993). Thus, there is a selfish incentive to not contribute to the provision of public goods and free-ride on the generosity of others. Evolutionary models have tried to explain the evolution of altruism and public good provision, and several models have invoked punishment of free-riders (e.g. Boyd, Gintis, Bowles, & Richerson, 2003; Boyd & Richerson, 1992; Fehr & Fischbacher, 2003; Gintis, 2000; Henrich & Boyd, 2000). Punishment can include physical aggression, public criticism, or ostracism, which tends to induce free-riders to cooperate. These authors have argued that punishment eliminates the selective advantage that defectors would otherwise have, thus increasing the prevalence of altruistic sentiments and cooperative behaviour.

However, most evolutionary models make no mention of which particular group members are more likely to provide public goods or punishment, and assume that altruism and punishment are equally likely and effective by any member of a population. Human groups regularly contain status differentials, such that some members have greater access to resources (including social resources) than others (Ellis, 1993; Hawley, 1999; Mazur, 1985). In industrial societies, researchers usually use socioeconomic status as a measure for individual status, but social status can encompass much more than that and include relative status within one's peer group. Such status differentials can include dominance based on fear of high status individuals or prestige based on respect (Henrich & Gil-White, 2001).

One might predict that high status individuals would contribute more to group efforts such as the provision of public goods than low status individuals if only high status people with access to money could afford to provide the public goods. However, there are other reasons to predict a relationship between status and the provision of public goods. Cooperation is good for all group members, but is individually risky because of the risk of exploitation, so a person's level of cooperation in repeated interactions depends on the likelihood of others cooperating (e.g. Dawes, McTavish, & Shaklee, 1977; Messick & Brewer, 1983; Smeesters, Warlop, Van Avermaet, Corneille, & Yzerbyt, 2003). People will often imitate cooperative actions of others (Fehr & Fischbacher, 2004), so a person's expectations of being imitated should also affect his/her level of

cooperation. Because persons of high status are likely to be listened to and imitated (Henrich & Gil-White, 2001), we can predict that they will be more likely to initiate collective action than low status persons. This can affect group cooperation, because the emergence of a group leader aids the cooperative conservation of common-pool resources in resource-sharing games (Muller & Vickers, 1996). Also, having status gives an individual confidence, which could make him/her more likely to undertake action and risk being a sucker in order to achieve a beneficial outcome. As an alternative hypothesis, high status people might be relatively uncooperative because they have the social power to free-ride upon others with relative impunity. Either way, these effects could occur even in anonymous laboratory environments if participants bring their outside preferences into the laboratory, or draw analogies between experimental games and real-life situations they are familiar with (Henrich et al., 2001).

Although some studies have investigated selection of leaders and authorities who alone have the ability to perform certain roles in social dilemmas (e.g. Van Vugt & De Cremer, 1999), there has been little work investigating the effects of status on cooperative behaviour when all group members have equal roles and behavioural options. Ball and Eckel and colleagues (1998; Ball, Eckel, Grossman, & Zame, 2001) found that people defer to high status participants in non-cooperative bargaining experiments and simulated markets. Barclay (*submitted*) found that self-reported perceptions of status were related to behaviour in a laboratory resource-sharing game. However, the direction of the

relationship was unclear because the particular resource-sharing game could have been construed as either a public good where resource-harvesting was cooperative or a common pool resource where resource-harvesting was uncooperative. Peterson, Ridley-Johnson, and Carter (1984) claimed that popular children were more likely to offer assistance to other children but their small sample size prevented statistical analysis. Ginsburg & Miller (1981) found that children who "altruistically" intervened in fights were likely to be socially dominant, and children who were rated as good leaders by their peers were more likely to sacrifice self-gain for their class or for needy children. However, these results have not been replicated with adults. Individuals who have higher experimental endowments tend to contribute more to non-linear public goods than other group members (Chan, Godby, Mestelman, & Muller, 1996, 1997; Chan, Mestelman, Moir, & Muller, 1996), although this is because the former have more to give and the particular public goods environment made it the Nash equilibrium for them to contribute more. Wealth was associated with giving more to a public good among the pastoral Orma in northern Kenya (Ensminger, 2004) and high status men among the horticulturalist Achuar in Ecuador tended to share meat with a higher number of people than low status men (Patton, 2004). However, these relationships did not hold in other small-scale societies (Henrich et al., *in press*). Finally, Kurzban and Houser (2001) found that people with high self esteem (as measured by the Rosenberg self-esteem scale) were less likely to be free-riders than people with low self-esteem, and self esteem could be a reflection of relative

social status. These studies provide suggestive evidence of a relationship between status and cooperation, but none have been experimental studies (with the exception of Ball & Eckel, which was non-cooperative bargaining), so it is unclear whether status affects cooperation or vice versa, and whether changes in status can increase cooperation.

Status is especially important to consider when examining the punishment of free-riders. Punishment from a low status individual towards a high status individual is likely to be ineffective if the low status person lacks the strength or social clout to harm the free rider either physically or socially without doing much more harm to himself. Punishment is especially dangerous when retaliation is possible, and retaliation can even occur in laboratory public goods games (Fehr & Gächter, 2000; Chapter 4 of this thesis). Punishment is much more likely to flow from a dominant individual towards a subordinate because a dominant has less to fear from defense and retaliation (Clutton-Brock & Parker, 1995), and this relationship presumably also holds in battles of reputation and social power. Having high status gives a person authority and legitimacy, so punishment might be more likely to be accepted if it comes from an authority or high status individual. In fact, punishing defectors may be a socially acceptable way to assert dominance, and theoretical models have predicted that high quality individuals may use punishment to signal their quality (Gintis, Smith, & Bowles, 2001). Based on these ideas, we would expect that high status makes a person more likely to punish others. Consistent with this, Wiessner (2003) found that high

status !Kung foragers are more likely to criticize the uncooperative actions of low status group members than vice versa. Among Zimbabwean villagers, more criticism came from members of households with high incomes, many marriage ties, and larger social networks than from members of households with lower incomes, fewer marriage ties, and smaller social networks (Barr & Kinsey, 2002). However, we cannot be certain whether status affects punishment or vice versa, or whether other variables (e.g. intelligence) affect both status and criticism.

The present study sought to test the effects of status on cooperative and punitive behaviour. In order to avoid confounds that might be associated with pre-existing status differences (e.g. intelligence), I attempted to manipulate the status of participants and then measured their contributions and punishment in a public goods game. Ball and Eckel (1998) manipulated status by giving participants false feedback on pre-game quizzes, and publicly congratulating people who allegedly did well. Similarly, Rutherford (*in press*) manipulated the outcome of tests in order to make some randomly selected participants feel superior to others. However, those status manipulations were either designed for two-person games and situations (Rutherford, *in* press) or were specific to particular fields (e.g. economics, Ball & Eckel, 1998), and I needed a status manipulation that could work on multiple participants from a variety of disciplines. Therefore, I created a quiz for use in this study. Participants received false information about how they ranked on the quiz, and they then played a public goods game with punishment. I

compared the contributions and punishment of people who allegedly won the highest rank with those who ranked lower on the false-feedback quiz.

3.2 Methods

3.2.1 Participants & Seating

Thirty-nine male (mean age: 18.4 years, S.D. 1.1 years) and fifty-eight female (mean age: 19.2 years, S.D. 1.3 years) undergraduate students from McMaster University participated. They received either one credit towards an introductory psychology course or $5 for their participation, in addition to any money they earned in the experiment (mean $5.06 ± S.D. $0.83). They participated in groups of four, and were seated at computers which were separated by plywood barriers (approximately 60 cm high) on top of desks. These barriers prevented participants from seeing each other's faces and computer screens but not from seeing the experimenter (2-3 m away, depending on seat location) or each other's legs and backs.

3.2.2 Status Manipulation

I attempted to manipulate status using a method based on Ball and Eckel (1998) and Rutherford (*in press*). Before playing the public goods game (see below), participants completed a quiz that purportedly measured "some of the skills that may be involved in the experimental game". This 20-question quiz started relatively easy and became progressively more difficult, such that many

participants could not completely finish the quiz within the 12 minutes given.[3] The difficulty of some questions and the time constraints ensured that no participant could be sure of his/her relative score. Participants received public feedback of their relative ranks (announced by codenames to maintain anonymity), and were told that their rank determined their player number in the public goods game. For example, the highest scorer "won the right to be Player 1" while the lowest scorer "got stuck with being Player 4" (even though player number actually conferred nothing special other than the name). In sessions with computerized players (see below), the rank of the free-riders and altruists was counterbalanced between sessions.

3.2.3 Public Goods Game

Participants played a Public Goods Game (PGG) with punishment (for details, see Fehr & Gächter, 2002) in single-sex groups of four, and were identified only by the player numbers assigned in the status manipulation. Participants earned "lab dollars" which were exchanged for Canadian dollars after the experiment at a rate of 10 to 1. Before each round of the public goods game, participants received 10 lab dollars. Each round, they could keep this money for themselves, or contribute any amount to a group fund. The experimenter multiplied the total contributions to the group fund by 1.6 before dividing this new total evenly among all participants. Thus, contributing was individually

[3] There were originally 25 questions, but some were removed after pilot testing because they were deemed too difficult to be credible questions. The time was also reduced to further challenge the participants.

costly, yet beneficial for the group, like a Prisoner's Dilemma with multiple players. After contributing, participants learned what each other participant had contributed and kept, and had the option of paying some of their earnings to punish the other players (of their choice) by reducing those persons' payoffs. Every dollar spent on punishment would reduce the punishee's payoff by three dollars. Participants would know how much they were punished, but not who did the punishing. After the punishment option, a new round began. There were five rounds, although participants did not know exactly how many rounds to expect. Data from three participants were excluded (before being examined) because it was apparent that they had not understood the instructions due to language difficulties, and from one participant because she indicated a belief that she was not actually playing the PGG with other participants.

In 20 groups (9 male, 11 female), participants were given false feedback about what the others in the group had contributed. This was done to attempt to standardize participants' experiences and to guarantee that each participant encountered a free-rider that he/she could punish. In these groups, each participant faced three different computerized players: a free-riding player who contributed \$1, \$1, \$0.5, \$0.5, and \$0 in the five rounds; an altruistic player who contributed \$9, \$9, \$8, \$7 and \$7; and a moderate player who contributed \$5, \$4.5, \$4, \$4, and \$3.5. The contributions of the computer players dropped each round because this is what normally happens in PGGs (Davis & Holt, 1993; Barclay, 2004). The computer players also responded to being punished by

increasing their contributions by one dollar for every dollar spent to punish them, and punished the participants for contributing very low amounts. This was done to make the game more realistic, because people are often punished for contributing low amounts and respond to punishment by increasing their contributions (Fehr & Gächter, 2002). There is no reason to expect this to cause certain subjects (higher or lower ranked) to contribute and punish more than others. If only three participants showed up to a session, a confederate replaced the absent participant, and was assigned the 2nd or 3rd rank in order to maximize the number of highest and lowest ranked participants. However, this resulted in uneven numbers of participants in each of the four status ranks.

Some participants indicated in a pre-PGG demographics questionnaire that they knew one of the other three participants in their session. Groups in those sessions played the PGG with each other instead of against the computer players because friends would be likely to conclude that there was false feedback if they talked after the game. These nine sessions (3 male, 6 female) with naturally occurring variation were not of primary interest, but were also analyzed in order to compare with the experimental games. In four groups where a real participant was absent, participants were still guaranteed exposure to a free-rider because the confederate filling in acted like a free-rider (contributed $1, $1, $0.5, $0.5, and $0 in the five rounds) who did not punish.

3.2.4 Questionnaires

After the PGG, participants completed two questionnaires so that I could estimate the effects of the manipulation on self-esteem and the relationship between self-esteem and tendencies to contribute or punish. Participants completed Rosenberg's Self-Esteem Scale (1965) measuring self-reported self-esteem, and Allen and Gilbert's (1995) Social Comparison Scale (SCS), which measures how subjects feel in relation to others on dichotomous characteristics such as superior-inferior (a subscale measuring Social Rank) or same-different (a subscale measuring Social Difference from one's group).

3.2.5 Data Analysis

In sessions with computerized players, sex and quiz ranking were analyzed as Between-Subjects factors in a 2 X 4 General Linear Model (SPSS 12.0). However, when participants played the PGG with each other instead of computer players, each group constitutes one unit of analysis because participants' decisions influence one another. In those sessions, there were only four groups with usable data in all status ranks because of the presence of a confederate or because a participant did not understand English (see above). There were seven sessions without computer players in which data were available for both the highest and lowest ranked participants, and they could be analyzed using a Repeated Measures General Linear Model (SPSS 12.0). This analysis reduces the number of males to two, so sex differences could not be computed for these sessions.

3.3 Results

3.3.1 Assigned Rank

In sessions with computerized players, feedback of one's ranking in the quiz had no discernible effect on first round PGG contributions or on overall PGG contributions (both Fs < 1, Figures 3.1 & 3.2). There was no significant effect of quiz ranking on first round PGG punishment or overall punishment ($F_{3,59}$ = 1.73 and 0.46, respectively, both n.s., Figures 3.3 & 3.4). Although men did not contribute more than females did in either the first or overall contributions (both Fs ≤ 1), they spent more on punishment in the first round and overall ($F_{1,59}$ = 4.42 and 7.11, both ps < 0.05, Table 3.1). Participant sex did not interact with quiz ranking on PGG contributions or punishment in either the first round or overall (all Fs < 1.2, all n.s.)

In sessions without computerized players, quiz ranking had no effect on first round PGG contributions, overall PGG contributions, first round punishment, or overall punishment ($F_{1,6}$ = 0.74, 4.57, 0.68, 2.05, respectively, all n.s., Figures 3.1, 3.2, 3.3, & 3.4). In overall PGG contributions, there was a trend for the lowest-ranked participants to contribute more (Mean = 27.36 S.E. = 5.20) than the highest-ranked participants (Mean = 23.71, S.E. 4.46), but this did not reach significance ($F_{1,6}$ = 4.57, p = 0.076), possibly because there were only seven observations. There were only two males in this analysis, so sex differences were not analyzed.

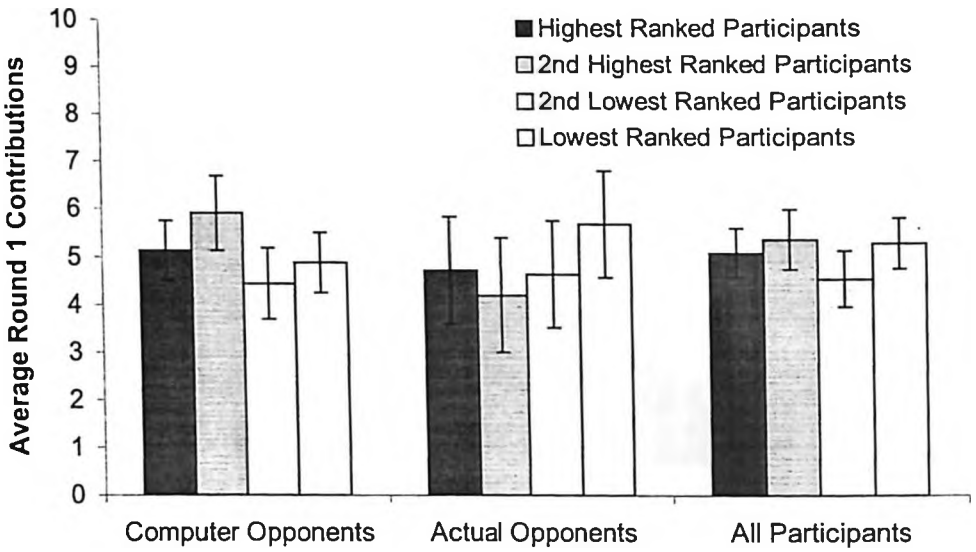


Figure 3.1: Effects of quiz rankings on first round PGG contributions (and standard errors of the means) for participants facing computer opponents (i.e. false feedback), actual opponents (i.e. no false feedback), and all participants combined. Note: after excluding confederates, there were relatively few participants facing actual opponents, with only six participants ranked second highest and eight participants in the other ranks.

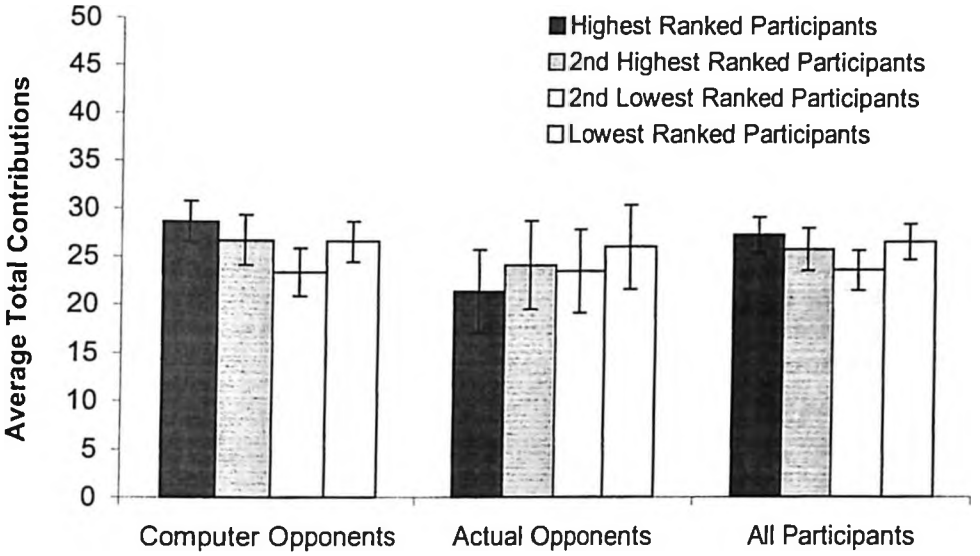


Figure 3.2: Effects of quiz rankings on total PGG contributions (and standard errors of the means) for participants facing computer opponents (i.e. false feedback), actual opponents (i.e. no false feedback), and all participants combined. Note: after excluding confederates, there were relatively few participants facing actual opponents, with only six participants ranked second highest and eight participants in the other ranks.

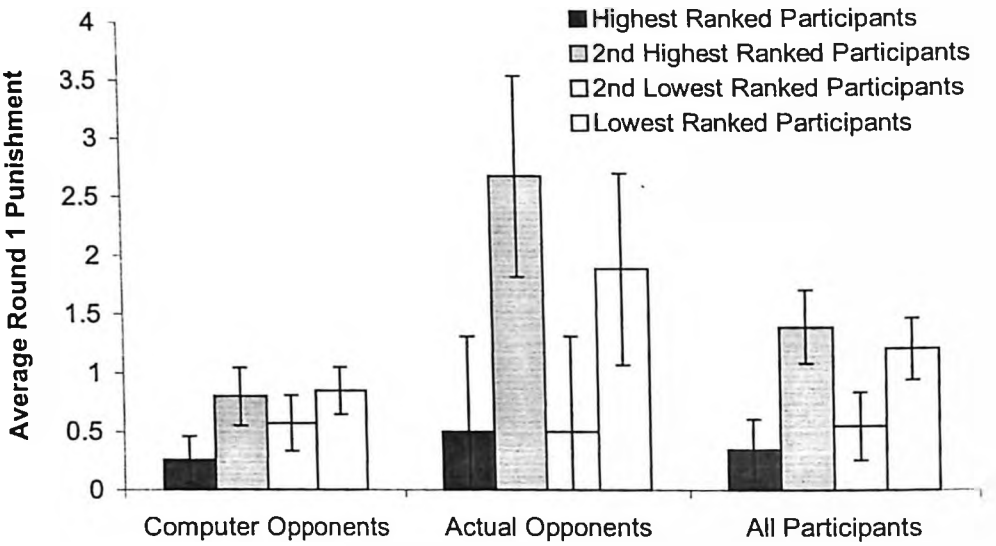


Figure 3.3: Effects of quiz rankings on first round PGG punishment (and standard errors of the means) for participants facing computer opponents (i.e. false feedback), actual opponents (i.e. no false feedback), and all participants combined. Note: after excluding confederates, there were relatively few participants facing actual opponents, with only six participants ranked second highest and eight participants in the other ranks. One participant who ranked second highest and faced actual opponents spent $9 on first round punishment; the abnormally high mean in that condition becomes more similar to other ranks (Mean = $1.58, S.E. = $0.58) when this outlier is excluded.

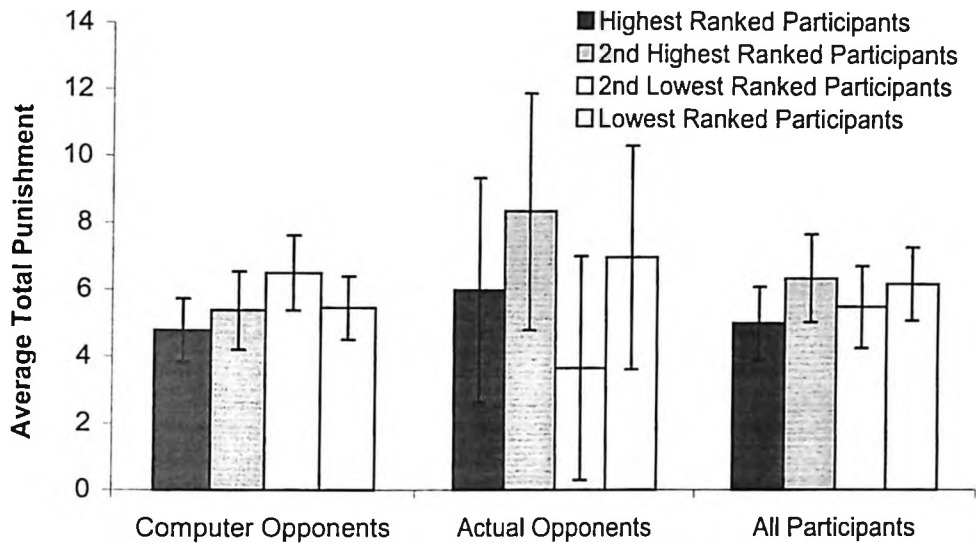


Figure 3.4: Effects of quiz rankings on total PGG punishment (and standard errors of the means) for participants facing computer opponents (i.e. false feedback), actual opponents (i.e. no false feedback), and all participants combined. Note: after excluding confederates, there were relatively few participants facing actual opponents, with only six participants ranked second highest and eight participants in the other ranks.

Table 3.1: Sex differences in PGG contributions, punishment, and self-esteem (and standard errors of the means). In sessions with computerized opponents, there were significant sex differences in first-round punishment, total punishment, self-esteem and perceived social status[4], but there were not enough males in the sessions without computer players to analyze sex differences.

	Computerized Opponents (31 males, 36 females)	Computerized + Real Opponents (39 males, 58 females)
Round 1 Contributions		
Male	5.21 (0.51)	5.14 (0.44)
Female	4.89 (0.48)	4.97 (0.36)
Total Contributions		
Male	27.79 (1.71)	26.47 (1.56)
Female	25.31 (1.60)	25.64 (1.29)
Round 1 Punishment		
Male	0.84 (0.16) *	1.03 (0.21)
Female	0.36 (0.15) *	0.66 (0.18)
Total Punishment		
Male	6.90 (0.77) **	6.69 (0.94) †
Female	3.97 (0.72) **	4.79 (0.77) †
Rosenberg Self-Esteem		
Male	30.42 (1.24) †	30.82 (1.11) ***
Female	27.19 (0.17) †	25.84 (0.91) ***
Social Comparison Scale (SCS)		
Male	76.32 (2.10) *	76.60 (1.96) ****
Female	69.83 (1.97) *	67.48 (1.62) ****

[4] Significance of sex difference: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$

I can combine the first round PGG contributions from sessions with and without computerized players, because participants' decisions have not yet affected one another's behaviour. Quiz ranking had no effect on first round contributions (Figure 3.1), nor did sex (Table 3.1), or the interaction between the two (all Fs < 1), but did have an effect on first round punishment ($F_{3,88} = 3.34$, $p < 0.05$, Figure 3.3) after controlling for others' first round contributions[5] (which approached significance as a covariate, $F_{1,88} = 2.81$, $p = 0.097$). However, the pattern is difficult to interpret because an analysis of orthogonal contrasts revealed a significant cubic relationship between rank and first round punishment ($p = 0.011$), but no linear or quadratic relationship (both n.s.). Highest-ranked participants spent \$0.32 (S.E. = \$0.25) on punishment, second-ranked participants spent \$1.37 (S.E. = \$0.30), third-ranked participants spent \$0.55 (S.E. = \$0.28), and lowest-ranked participants spent \$0.96 (S.E. = \$0.26). Sex had no effect on first round contributions or punishment, nor did it interact with quiz rankings (all Fs < 1, Table 3.1). I cannot correctly analyze total contributions or punishment in all sessions combined, even if I control for the contributions of others and the amount of punishment received, because participants will have affected each other's behaviour by the end. If I try, quiz ranking has no effect on either total contributions or total punishment (Fs$_{3,87}$ = 1.01 and 0.28, both n.s., Figures 3.3 & 3.4). Although males did not contribute more overall than females ($F < 1$, n.s.),

[5] The results were the same if I control for the lowest contribution in the group, or the difference between each participant's contributions and the group average or minimum.

they spent slightly more overall on punishment, although this was not quite significant ($F_{1,87} = 2.74$, $p = 0.10$, Table 3.1).

There was no relationship between quiz rankings and scores on either the Rosenberg Self-Esteem Scale or the Social Comparison Scale (Table 3.2) in participants who played against computerized players (both Fs < 1), participants who played with real players (both Fs < 1), or all participants combined (both Fs < 1). These results were the same whether I used the entire Social Comparison Scale or only the Social Rank subscale. Males scored higher than females on the Rosenberg Self-Esteem Scale and the Social Comparison Scale in all three analyses (computerized players: $Fs_{1,59} = 3.07$ and 4.11, ps = 0.085 and 0.047, respectively; real players: $Fs_{1,21} = 8.78$ and 5.98, ps = 0.007 and 0.027, respectively; all participants: $Fs_{1,89} = 11.34$ and 12.74, respectively, both ps < 0.005; Table 3.1). Among participants who played computerized players, scores on the Rosenberg Self-Esteem Scale and Social Comparison Scale had no relation to round 1 contributions, total contributions, or round 1 punishment ($rs_{67} < 0.20$, all ps > 0.1), but they both had a positive correlation with total punishment ($rs_{67} =$ 0.25 and 0.27, respectively, both ps < 0.05). However, there were no significant correlations among participants who played with real players (all $rs_{30} < 0.25$, all ps > 0.2) or in all participants combined (all $rs_{97} < 0.15$, all ps > 0.15, possibly because of the noise introduced by different amounts of contribution and punishment.

Table 3.2: Effects of quiz rankings on scores on the Rosenberg Self Esteem and Social Comparison Scales (and standard errors of the means). None of these differences are statistically significant.

	Computerized Opponents (67 participants)	Real Opponents (30 participants)	Both Combined (97 participants)
Rosenberg Self Esteem			
Highest-Ranked	28.05 (1.53)	25.25 (2.72)	27.25 (1.32)
2nd Highest-Ranked	30.23 (1.90)	25.83 (2.89)	28.84 (1.58)
2nd Lowest-Ranked	27.57 (1.82)	26.88 (2.72)	27.32 (1.48)
Lowest Ranked	29.10 (1.53)	25.14 (3.60)	28.07 (1.36)
Social Comparison Scale			
Highest-Ranked	72.10 (2.59)	63.38 (5.43)	69.61 (2.34)
2nd Highest-Ranked	74.46 (3.21)	73.33 (5.76)	74.11 (2.81)
2nd Lowest-Ranked	74.21 (3.08)	68.74 (5.43)	72.23 (2.62)
Lowest Ranked	71.55 (2.59)	66.00 (7.18)	70.11 (2.41)

3.3.2 Proximity

Because the computers for the experiment were positioned in a square, two participants in each session were seated closer to the experimenter than the other two. This physical proximity was counterbalanced across the assigned quiz rankings, such that high ranking participants sat close to the experimenter as often as low ranking participants. However, proximity seemed to have significant effects on participant behaviour. For simplicity, the following effects are presented for participants who played against computers only, but the results are very similar (see Table 3.3) if I include all subjects. Participants who were seated close to the experimenter contributed less in the first PGG round than participants who were farther away ($F_{1,63} = 5.98$, $p = 0.018$), although they did not contribute

Table 3.3: Average (and standard errors of the mean) cooperation and punishment (in lab dollars) in relation to physical proximity to experimenter[6].

	Computerized Opponents (67 participants)	Real Opponents (30 participants)	Both Combined (97 participants)
Round 1 Contributions			
Close to Experimenter	4.31 (0.45) *	4.94 (0.62)	4.52 (0.37) *
Far from Experimenter	5.79 (0.45) *	5.19 (0.71)	5.62 (0.39) *
Total Contributions			
Close to Experimenter	25.38 (1.58)	25.53 (2.46)	25.43 (1.35)
Far from Experimenter	27.56 (1.60)	24.08 (2.82)	25.58 (1.43)
Round 1 Punishment			
Close to Experimenter	0.47 (0.16)	0.82 (0.47)	0.59 (0.19)
Far from Experimenter	0.70 (0.16)	1.92 (0.54)	1.04 (0.21)
Total Punishment			
Close to Experimenter	5.18 (0.69)[a]	3.88 (1.87) †	4.74 (0.79) [a]
Far from Experimenter	5.48 (0.71)[a]	8.92 (2.13) †	6.46 (0.83) [a]

more overall ($F_{1,63} = 1.11, p = 0.29$). There was no difference in first round punishments ($F_{1,63} = 1.17, p = 0.28$), although there was a marginally significant Sex X Physical Proximity interaction in overall punishment ($F_{1,63} = 3.92, p = 0.052$). Proximity had no effect on men's overall punishment (Near: Mean = \$7.75, S.E. = \$1.23; Far: Mean = \$6.60, S.E. = \$1.27; $F < 1$, n.s.), but women tended to punish more if they sat farther from the experimenter (Mean = \$5.06, S.E. = \$0.73) than if they sat closer (Mean = \$2.89, S.E. \$0.7) ($F_{1,34} = 4.45, p = 0.045$). Including proximity as a covariate in the analyses of quiz ranking does not change any of the findings.

[6] † $p < 0.10$, * $p < 0.05$, a: Sex X Proximity interaction with $p < 0.10$

3.4 Discussion

Assigned rank had no apparent effect on either PGG contributions or punishment, either in the first round or overall. The only significant effect involving rank was that second-ranked players punished more in the first round than other participants in a combined analysis of all participants. However, men punished more than women did, and high self esteem (as measured by the Rosenberg Self-Esteem Scale and Social Comparison Scale) was related to high levels of overall punishment. Finally, physical proximity to the experimenter was related to high initial contributions and total punishment.

The particular relationship between rank and first-round punishment was rather surprising, and not readily interpretable because second-highest ranked participants punished the most, then the lowest-ranked, then third-highest-ranked, and the highest-ranked punished the least. Combining all participants into one analysis like that is technically not correct but was presented solely to show that combining all sessions does not add to the power and enable small effects to be found. This pattern is not similar to any other effects, and there was absolutely no effect for overall punishment. It is possible that the effect in first round punishment is a spurious effect caused by multiple comparisons. If so, then it would appear that assigned rank had no real effect on any measure in the PGG. This lack of effect is unlikely to have been caused by low statistical power, because there was enough power to detect that men punished significantly more

than women in most analyses, whereas most effects involving assigned rank did not even approach significance.

It is possible that the null hypothesis is true, and assigned status has no effect on cooperation or punishment. A second possibility is that different participants responded differently to the status manipulation. Utz, Ouwerkerk, & Van Lange (2004) found that "prosocial" participants (as rated by a prior test of social values) increased their levels of cooperation in a social dilemma after being primed with words related to competence and ability, whereas this priming caused "competitive" participants to decrease their cooperation. If a similar effect occurred in this study, then the prosocial high-ranked participants would increase their cooperation in response to the status manipulation, and the competitive high-ranked participants would decrease their cooperation, and these effects would cancel out to leave no overall effect of rank.

A third possible reason for the lack of effect is that the status manipulation was ineffective. Assigned rank had no effect on self-esteem scores as measured by either the Rosenberg Self-Esteem Scale or the Social Comparison Scale, and this was probably not due to low statistical power because men had significantly higher self-esteem than women on both scales. If assigned rank affected perceptions of status, then it presumably would have affected scores on at least one of these scales. Given that there was no perceivable effect of assigned rank on anything that I measured, it is likely that the manipulation was not successful. Anecdotal evidence for this also comes from the debriefing of participants. While

being told after the experiment (for ethical reasons) that the quiz rankings were not meant to reflect any underlying quality or worth and that participants might have ranked differently in different groups, many participants appeared unconcerned with their rank, and several made explicit comments that they were unconcerned. Unfortunately, there is no record of which participants these were so they cannot be excluded from the analyses, but the comments do provide further support for this explanation.

I do not conclude that status does not affect cooperation or punishment. Participants' levels of punishment were positively related to their self esteem scores as measured by the Rosenberg Self Esteem scale and the Social Comparison scale. Also, self esteem and perceptions of social rank were related to cooperative behaviour in social dilemmas in previous studies (Barclay, *in preparation*; Kurzban & Houser, 2001). The results of the present study need to be replicated given that self-esteem did not predict PGG contributions. Exposure to a strong free-rider (the computerized strategy) could have crowded out any effects of self esteem on cooperative behaviour, just as situational variables often account for more cooperation than personality factors in social psychological research (Liberman, Samuels, & Ross, 2004; Myers & Spencer, 2001). This situational factor would not overwhelm sanctioning behaviour, because the presence of a free-rider is exactly the type of situation that brings out punitiveness.

The relationship between self esteem and punishment warrants further investigation. This could be done with a more effective status manipulation, possibly involving role-playing of high or low status roles (Snodgrass, 1985, 1992) and two-player cooperative games. Because the legitimacy of leaders is important in social dilemmas (Van Vugt & De Cremer, 1999), it is possible that status only affects cooperation and punishment if all parties involved acknowledge the existence of the status differences, such that only pre-existing and relatively stable differences have an effect. Perhaps cooperative games could be played with people who have just won or lost contests or games (or observed a win or loss), because winning and losing contests have been linked to rising or falling testosterone levels in males that may be associated with status changes (Booth, Shelley, Mazur, Tharp, & Kittok, 1989; Mazur & Booth, 1998). It might even be possible to have participants reflect on a time when they rose or fell in status, because imagined success has also been shown to have an effect on testosterone levels (Schultheiss, Campbell, & McClelland, 1999), and inducing guilt or anger this way can affect cooperation (Ketelaar, 2003) or the interpretation of instances of cheating (Chang & Wilson, 2004).

It was interesting that participants who sat closer to the experimenter behaved slightly differently than participants who sat farther away. Despite the absence of eye-to-eye contact between participant and experimenter, physical proximity could have induced demand characteristics or made the participants want to make a more socially acceptable decision. Physical proximity and degrees

of anonymity have effects on obedience (reviewed by Myers & Spencer, 2001), so this finding could extend the effects of proximity to cooperative behaviour. Anonymity normally increases altruistic behaviour in cooperative games (Gächter & Fehr, 1999; Rege & Telle, 2004), and even one person (the experimenter) knowing one's decisions is enough to increase altruism (Hoffman, McCabe, Shachat, & Smith, 1994) or punishment (Kurzban, 2004). Physical proximity could affect cooperative behaviour by reducing perceptions of anonymity or subtly triggering reputational concerns, which have recently been shown to affect levels of cooperation (Burnham & Hare, *in press*; Haley & Fessler, 2005). Alternatively, physical proximity to the experimenter could induce demand characteristics such that participants do what they believe is normative behaviour or instead what is "rational" (in this case, spending less on contributions and punishment), and these explanations are not mutually exclusive.

The effects of feedback about others' contributions could overwhelm any effects of proximity just as other situational variables overwhelm personality variables in social psychological research (Myers & Spencer, 2001), so it is not surprising that proximity is related to first round PGG contributions but not overall contributions. With punishment in the first round, people are probably unsure about whether punishing free-riders is socially acceptable (see Chapter 4) so proximity could have mixed effects on first round punishment. With repeated rounds, it becomes easier to guess the motives of repeated free-riders and therefore more socially acceptable to punish them, so any effects of proximity

would be expected to arise in later-round or overall punishment, which they did. Of course, these hypotheses are post hoc explanations of an unanticipated finding, and require further investigation before any firm conclusions can be drawn, but greater efforts should be used to control for physical proximity in order to reduce the statistical noise that it might add to future experiments. Fortunately, my subsequent studies (see Chapter 4) have had more effective visual barriers that prevented participants from seeing any body parts of the experimenter or other participants rather than just preventing eye contact and sight of each others' computer screens.

While this study failed to find any evidence that cooperation and punishment are affected by manipulated status, it did suggest that they could be affected by self-esteem and physical proximity to the experimenter. Also, it provided some participants with exposure to a repeated free-rider, which was useful for a questionnaire that participants filled out after completing this study. In that questionnaire, participants imagined that they were in a group with a person who punished free-riders and a person who did not punish, and then rated both the punisher and non-punisher on nicety, trustworthiness, group-focus, and worthiness of respect. This questionnaire was a pilot for the following set of studies regarding the trustworthiness of altruistic punishers, and the results are presented as Study 1 of the following chapter. Thus, although the results of this study were disappointing, they led to some important questions for further investigation.

Chapter 4: Altruistic punishment and reputation: Is it advantageous to punish free-riders?

Abstract

Many studies show that people act cooperatively and are willing to punish free-riders (i.e. people who do not cooperate). However, non-punishers benefit when free-riders are punished, making punishment a group-beneficial act. I report five studies investigating whether punishers gain social benefits from punishing. Undergraduate participants played public goods games (cooperative group games involving money) in which there were free-riders, and in which they were given the opportunity to impose monetary penalties on free-riders. Participants rated punishers as being more trustworthy, group-focused and worthy of respect than non-punishers. In dyadic trust and cooperation games following public goods games, punishers did not receive monetary benefits from punishing free-riders in a single-round public goods game, but did benefit monetarily from punishing free-riders in a repeated public goods game. Punishment that was not directed at free-riders brought no monetary benefits, suggesting that people distinguish between justified and unjustified punishment and only respond to punishment with enhanced trust when the punishment is justified.

4.1 Introduction

In order for altruism among unrelated individuals to evolve, altruists must be able to identify non-altruists and defectors (Cosmides & Tooby, 1992), and either punish them or avoid them (see for example, Axelrod, 1984). This is particularly difficult when altruism cannot be directed towards specific individuals, such as in the provision of public goods. Public goods are things that people have to expend time, effort, or money to provide, but once they are provided, others cannot readily be excluded from benefiting even if they did not contribute the provision of the public good (Davis & Holt, 1993; Messick & Brewer, 1983). Classic examples of public goods include irrigation, group protection and vigilance, or any collective action project. Public goods are collectively beneficial, but free-riders who cooperate relatively little are better off than cooperators, causing selection for non-cooperation that should eventually undermine collective action. Consistent with this, cooperation in laboratory experiments drops when group members find out that others have contributed less than themselves to the provision of public goods (e.g. Andreoni, 1995; Fehr & Fischbacher, 2004a), presumably because contributors retaliate by also contributing less (Gintis, 2000).

The opportunity to impose sanctions on free-riders can potentially solve this collective action problem and allow for the evolution of cooperation because being punished induces free-riders to cooperate (e.g. Boyd & Richerson, 1992; Caldwell, 1976). In non-laboratory settings, such sanctions can include criticism,

ostracism, and physical or social threats, all of which carry risks of retaliation, enmity, or the loss of partnership. In typical laboratory experiments, the punisher has to pay a monetary cost to reduce the payoff of other players. Despite these costs, it is clear that some people will punish free-riders when they have that option in laboratory experiments (e.g. Fehr & Gächter, 2002; Ostrom, Walker & Gardner, 1992; Yamagishi, 1986) and in field settings (Barr, 2001; Cordell & McKean, 1992; Price, 2005), and that this raises the level of cooperation in groups.

In cooperative group situations, punishing a free-rider can be considered an altruistic[7] act towards other group members because everyone benefits from the resulting increase in the free-rider's level of cooperation (Yamagishi, 1986). People without punitive sentiments might be expected to benefit from punishment opportunities more than people who have punitive sentiments and act on them, because the former do not pay the cost of imposing sanctions and yet still benefit from the punishment provided by the latter. If this occurred in ancestral environments, then punitive sentiments should have been selected against. Punishment could also decrease in frequency within an individual's lifetime if he/she learns (from experience and observation of others) that punishing brings fewer relative gains than not punishing. People should notice and care that non-punishers are better off than punishers given that humans care about their payoffs

[7] I am using a definition of altruism that focuses on the average immediate costs to the actor and benefits to the recipient, not the underlying psychological motivations (e.g. Trivers, 1971). By this definition, acts that help someone else at Time A and are (on average) individually costly are considered altruistic even if they bring social benefits or induce future reciprocation at Time B.

relative to others (e.g. Bolton & Ockenfels, 2000; Roth, 1995), are sensitive to people taking benefits without paying the appropriate costs (Cosmides & Tooby, 1992), and can learn by observation (Tomasello, Kruger & Ratner, 1993). Thus, punishment should decrease in frequency both within generations (via learning) and across generations (as punitive sentiments are selected against) unless there is some process by which punishment itself is rewarded.

Some game theoretic models and computer simulations of the evolution of punishment postulate that punishers benefit from being in cooperative groups. Groups with sanctions will have higher cooperation than groups without, so the former will tend to outcompete the latter and will be less likely to disband. The between-group advantage of having sanctions in a group (and consequently higher cooperation) can be greater than the within-group disadvantage that punishers face, so that the level of altruism and altruistic punishment will tend to increase in the population (Boyd et al., 2003; Gintis, 2000; Sober & Wilson, 1998). Once punishers become common, there is less need to punish because free-riding will be rare, so there is not a big difference between the payoffs to punishers and non-punishers. Punishment can then be maintained in a group by a weak tendency to imitate the behaviour of others (conformist transmission, Henrich & Boyd, 2001). Punishment (and other group-beneficial norms) can spread between populations when less successful groups imitate the norms of cooperative yet punitive (and hence, more successful) groups (Boyd & Richerson, 2002). However, these models are unclear on how punishment becomes common within groups in the

first place if punishers are disadvantaged relative to non-punishers and non-punishment is the socially prescribed (and hence, most common) behaviour.

If punishers receive personal benefits for their punitive behaviour that other group members do not gain, then people could learn to punish. If this also occurred in ancestral environments, then natural selection could have favoured the punitive sentiments that motivate punishment. When punishment is group-beneficial, punishers may receive the same type of reputational benefits that altruists receive for their altruism, such as rewards from others. Three studies gave participants an opportunity to make donations to other participants (Wedekind & Milinski, 2000), to a group fund (Milinski, Semmann, & Krambeck, 2002a, b) or to charity (Milinski et al., 2002b), and found that the people who were the most generous were most likely to receive donations from others even though direct reciprocation of generosity was not possible. Similar rewarding of altruists has also been found among the Ache hunter-gatherers of Paraguay (Gurven et al., 2000). It remains to be seen whether people will reward punishers. Altruism could also signal trustworthiness, in that altruists are expected to be less likely to cheat in cooperative partnerships (Alexander, 1987). Barclay (2004; Chapter 2) found that people who made high contributions in a cooperative group game were trusted with more money in a subsequent dyadic trust game than those who made lower contributions. When punishing a free-rider is good for a group, it could signal the punisher's trustworthiness, commitment to that group, concern with fairness, or unwillingness to tolerate being cheated. This signal need not be a

conscious one; it can function as a signal as long as people respond in certain ways to those who display punitive sentiments.

If punishment is a signal of trustworthiness or fairness (for example), then punishers may receive benefits from others who are acting solely out of self-interest. Others might be more willing to enter and invest more in relationships with people who have demonstrated that they will not tolerate unfairness, such that punishers receive more benefits from cooperative partnerships than non-punishers. Being known for imposing sanctions could be beneficial if other people are less likely to cheat on sanctioners out of fear of retaliation (Brandt, Hauert & Sigmund, 2003). Although punishing non-punishers and rewarding punishers are altruistic acts that would require explanation themselves (Henrich & Boyd, 2001), trusting and fearing punishers are not subject to this "second-order" sanctioning problem. It can be in an observer's best interest to enter cooperative relationships with punishers in order to gain a trustworthy partner, and avoid cheating in those relationships in order to avoid sanctions. Thus, if there are reputational benefits for punishing, trust and respect (or fear) are likely candidates.

The present set of studies tests the hypothesis that punishers receive reputational benefits for sanctioning free-riders. Currently, there are no empirical studies bearing on this hypothesis. The alternative hypotheses are that punishers acquire a bad reputation because of the negative nature of sanctions, or that punishing does not lead to reputational consequences. Study 1 examined people's attitudes towards people who punished free-riders, and Studies 2-5 tested whether

punishers actually received more monetary benefits in experimental trust and gift exchange games than non-punishers.

4.2 General Methods For Public Goods Game

Undergraduate participants from McMaster University were recruited from an introductory psychology course (in exchange for course credit), and played a cooperative group game known as a Public Goods Game (PGG) with punishment (for details, see Fehr & Gächter, 2002) in groups of four. Each participant was given a pseudonym so that he/she could acquire a reputation in the game yet still remain anonymous. Participants earned "lab dollars" which were exchanged for Canadian dollars after the experiment at a rate of 10 to 1. Before each round of the PGG, participants received 10 lab dollars. In each round, they could keep this money for themselves, or contribute any amount to a group fund (the public good). The experimenter multiplied the total contributions to the group fund by 1.6 before dividing this new total evenly among the four participants. Thus, contributing was individually costly, yet beneficial for the group, like a Prisoner's Dilemma with multiple players. After contributing, participants found out what each other participant had contributed and kept, and had the option of paying some of their earnings to punish other participants (of their choice) by reducing those persons' payoffs. Unless otherwise noted, every dollar spent on punishment would reduce the punishee's payoff by three dollars, and all players were informed after each round of who had punished whom. After the punishment

option, a new round began. Participants did not know exactly how many rounds to expect.

4.3 Study 1

Study 1 gave people experience in a cooperative game with a conspicuous free-rider. Afterwards, participants gave their views of people who punish free-riders and of people who do not. Because of the negative nature of sanctions, punishers will not necessarily be liked more than non-punishers. However, if punishment can signal prosocial qualities like trustworthiness or commitment to a group, then punishers should be deemed more trustworthy, group-focused and worthy of respect than non-punishers. It is possible that only people who are punishers themselves will interpret others' punishment as a signal of prosocial qualities, so I also tested whether or not people's ratings of punishers (relative to non-punishers) was related to their own punitive behaviour.

4.3.1 Study 1 Methods

4.3.1(1) Public Goods Game:

Thirty male (average age 18.5 ± 1.1 years) and twenty-two female (average age 18.9 ± 0.9 years) undergraduate students played a Public Goods Game (PGG) with punishment for five rounds. Unbeknownst to the participants, they were not actually playing the PGG against the others in the room, but against computer players. One of these computer players was programmed to behave selfishly and contribute 1, 1, 0.5, 0.5 and 0 dollars in each of the five rounds. The

other two computer players were programmed to behave cooperatively (9, 9, 8, 7, & 7 dollars) or relatively neutrally (5, 4.5, 4, 4, & 3.5 dollars). The contributions of the computer players were designed to drop each round because this is what normally happens in PGGs (Davis & Holt, 1993). The computer players were also designed to respond to being punished by increasing their contributions by one dollar for every dollar spent to punish them, and to punish the participants for contributing very low amounts. This was done to make the game more realistic, because people often receive punishment if they contribute low amounts, and often respond to punishment by increasing their contributions (Fehr & Gächter, 2002). Participants were told how much they were punished, but not who did the punishing.

4.3.1(2) Measuring Feelings Towards Punishers and Non-punishers:

After having the experience of playing the PGG, participants filled out a questionnaire asking them to imagine the situation in which there was a person in their group who did not contribute anything to the group. Participants then rated how they would feel about someone who punished a non-contributor using 7-point Likert scales with anchors of mean/nice, untrustworthy/trustworthy, self-focused/group-focused, and unworthy/worthy of respect. Using the same scales, participants also rated how they would feel about someone who did not punish the non-contributor. The data were analyzed using a Repeated-Measures General Linear Model on SPSS (Version 11.0) comparing feelings about punishers to feelings about non-punishers, and participant sex.

4.3.2(3) Context:

This study was conducted in conjunction with an experiment (Chapter 3) testing for effects of manipulated status and self-esteem on levels of cooperation and punishment. Prior to playing the experimental cooperative game, participants were given a difficult skill-testing quiz and were given false feedback that they performed well or poorly on the quiz relative to the others in the group. As this feedback had no significant effect on levels of cooperation, punishment, or self-esteem (measured after the game), and has already been discussed in Chapter 3, it will not be mentioned further.

4.3.2 Study 1 Results

At least 25% of participants punished in each round, and 88% of participants punished at least once. Men spent more on sanctions than did women (means: \$1.37/round vs. \$0.86, $F_{1,50} = 4.02$, $p = 0.05$). The majority (81%, 114/141) of the punishment decisions were directed solely at the free-riding computer strategy. High-contributing participants sometimes (6% of punishment instances, 9/141) punished both the moderate-contributing computer strategy and the free-rider, and participants who had received punishment in previous rounds sometimes punished cooperators and the free-rider (10/141 punishment instances, 7%) or cooperators alone (2%, 3/141). Only five instances of punishment (4%) had no obvious provocation.

Participants did not perceive the punishers as being significantly nicer than the non-punishers ($F < 1$), but they did feel that the punishers were more

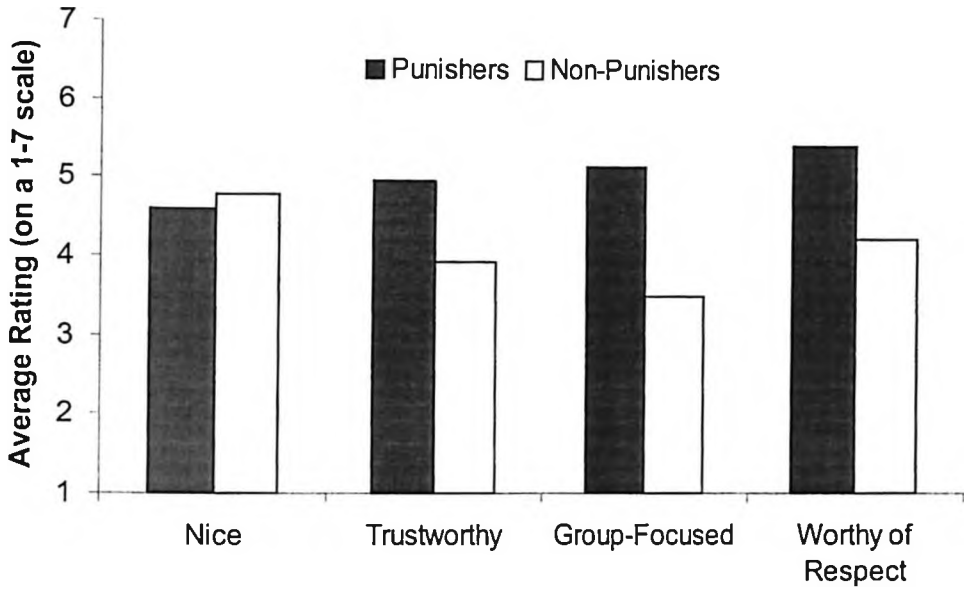


Figure 4.1: Average ratings on a 7-point Likert scale of feelings towards punishers (black bars) and non-punishers (white bars). Higher values represent more positive impressions.

trustworthy, group-focused, and worthy of respect ($Fs_{1,48} = 7.47, 13.80, 15.13$, respectively; all $ps < 0.01$, Figure 4.1). There was no interaction of participant sex with ratings on any of these characteristics ($Fs_{1,48} = 2.15, 1.01, 1.89$, and 0.02, respectively, all n.s.), nor did sex have any main effects (all $Fs < 1$). Participants who were punished in at least one round for low contributions (n = 38) did not differ from participants who received no punishment (n = 14) to the extent to which they thought that punishers were nicer, more trustworthy, more group-focused or more worthy of respect than non-punishers (all $Fs < 1.2$). Also, there were no interactions between sex and punishment received on the extent to which

participants perceived punishers differently than non-punishers on these characteristics ($Fs_{1,48}$ = 0.02, 0.69, 2.44, and 2.06, respectively, all n.s.).

There was a significant correlation between how much a participant spent on sanctions and the extent to which he or she thought punishers were nicer than non-punishers (r_{50} = 0.36, p = 0.009). However, this correlation was not significant for trustworthiness, group-focus, or worthiness of respect (rs_{50} = 0.16, 0.22, 0.21, ps = 0.27, 0.12, 0.14, respectively). This suggests that people's attitudes about punishers' trustworthiness, group-focus, or worthiness of respect are not simply byproducts of the amounts that they spent on punishment. Neither were these attitudes significantly predicted by individual contributions (rs_{50} = 0.13, 0.24, 0.17, ps = 0.36, 0.09, 0.41, respectively).

4.3.3 Study 1 Discussion

This study showed that after encountering a free-rider in a public goods game, people perceive punishers as being more trustworthy, group-focused, and worthy of respect than non-punishers. This was not due to a general positive impression of the punishers (a "halo effect"), because punishers were not seen as nicer than non-punishers. These perceptions were not affected by participant sex, nor by whether the participant received sanctions. Also, these results do not appear to have been caused by punishers favouring other punishers, because there was no significant correlation between individual punishing behaviour and the extent to which subjects thought punishers were more trustworthy, group-focused, and worthy of respect than non-punishers were.

4.4 Study 2

The results of Study 1 were suggestive, and could translate to benefits for the punishers if these views affect people's behaviour in non-laboratory environments. However, to make this claim, we need data on whether people will actually invest real money to trust or respect or reward people who have sanctioned free-riders, as they have been shown to trust and reward non-punitive altruists (Barclay, 2004, Chapter 2; Milinski et al., 2002). Study 2 attempted to show this by having participants play a PGG, and then play an experimental Trust Game with punishers and non-punishers to see whether they would trust punishers more than non-punishers.

4.4.1 Study 2 Methods

Twenty-two undergraduates (20 female, 2 male, average age 18.9 ± 0.9 years) played a PGG (with punishment). They were not told how many rounds they would play, and in fact they played only one round of the PGG. As in Study 1, the participants were led to believe that they were playing with the other participants in the room, but they were actually playing against pre-programmed computer players. One computer player was a free-rider (contributed $1), one was a punisher (contributed $7, spent $2 to punish the free-rider), and one was a non-punisher (contributed $7, did not punish).

After one round of the PGG, participants played a modified version of Berg, Dickhaut and McCabe's (1995) Trust Game (also known as the Investment Game) using the same pseudonyms that they had used in the PGG. In the Trust

Game, players were paired, and each received $10. One member of the pair (the Truster) had the option to send any number of these dollars to the other member (the Responder), and any amount sent got tripled.[8] The Responder could then return as much or as little of the tripled amount as he/she desired. Thus, Trusters could have increased their payoffs if they trusted Responders, and those Responders repaid that trust. To gather data on trust towards each of the other "players", I modified the game by using the "strategy method" (e.g. Fehr & Fischbacher, 2004b): participants indicated how much they wanted to entrust to each other player, and these decisions were elicited in random order. Participants were told that those decisions were binding because they would be randomly paired and assigned to roles within a pair and their corresponding trust decision would be implemented if they were assigned to be the Truster in their pair. Thus, the amount entrusted to different "players" in the Trust Game was a within-subject factor that was analyzed with a Repeated-Measures General Linear Model (SPSS 11.0). The "strategy method" was not used to measure Responder behaviour because participants would have had to indicate their preferred responses for each possible amount they could have received. Instead, all participants were told that they had been assigned to be the Responder, and were told which "player" they had been paired with (these pairings had been randomly assigned in advance). Thus, there were far fewer data points for that variable. Five participants were paired with the free-rider, six with the non-punisher, and ten

[8] During the experiment, these roles were referred to as the "First Mover" and "Second Mover".

with the punisher. Data on one participant's Responder behaviour were not available because of a computer error. They were then told that they had been sent $8, which indicates a reasonable level of trust, and were then asked how much of the tripled amount ($24) they wanted to return to the other "player".

4.4.2 Study 2 Results

Participants contributed an average of $5.36 (S.D = 2.82) in the PGG. Of the 22 participants, 14 (64%) punished one of the other "players" (the free-rider in 13/14 cases), and the average amount spent on sanctions was $0.91 (S.D. = 0.81). There was a positive correlation between how much a participant contributed in the PGG and how much he/she trusted the other three "players" in the Trust Game on average ($r_{20} = 0.47, p = 0.026$). However, there was no significant correlation between the amount that a participant spent on punishment and the amount he/she contributed in the PGG ($r_{20} = 0.04$, n.s.) or trusted the other three "players" in the Trust Game ($r_{20} = 0.05$, n.s.).

There were significant differences between the amounts entrusted to free-riders, punishers, and non-punishers ($F_{2,42} = 26.79, p < 0.001$, see Figure 4.2), and an analysis of orthogonal contrasts revealed that free-riders were trusted less than contributors (punishers and non-punishers) ($F_{1,21} = 27.32, p < 0.001$). However, punishers and non-punishers were not entrusted with different amounts ($F < 1$). In fact, 15 of the 22 participants sent exactly the same amount to the punisher and the non-punisher, and of the 7 who sent different amounts, 4 sent more to the

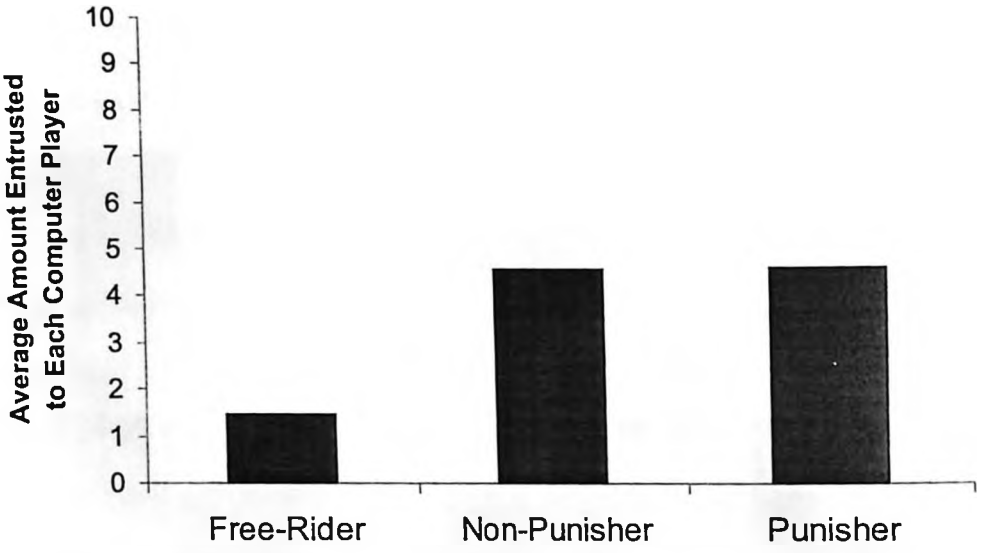


Figure 4.2: Average amounts entrusted to free-riders, non-punishers and punishers in the Trust Game after one round of a public goods game in Study 2. Free-riders received less than punishers and non-punishers, yet there were no differences between punishers and non-punishers.

punisher and 3 sent more to the non-punisher. There was no relationship between the amounts that participants contributed or punished in the PGG and how much they entrusted to punishers rather than non-punishers. There were not enough males to examine sex differences, nor were there enough data points to analyze Responder behaviour because the "strategy method" was not used.

4.4.3 Study 2 Discussion

Although free-riders were trusted less than contributors (replicating Barclay, 2004; Chapter 2), punishers were not trusted more than non-punishers in this study. This could be because of a couple of factors. The computer punisher and non-punisher both contributed the same amount in the PGG, and a desire to treat them equally based on their contributions could have overwhelmed the effect

of punishment on reputation when participants were asked what they would send to all three other "players". If participants had to decide what to send to only one other "player", then perhaps punishers would have been trusted more than non-punishers because there would be no direct comparison. Also, the Trust Game may not have been the best game to test for the effects of punishing on one's reputation. Study 3 was conducted to address these possibilities.

4.5 Study 3

Punishers might benefit from being rewarded by others, trusted by others, or feared by others. Berg et al.'s (1995) Trust Game could arguably measure the first two, but does not measure fear. For this reason, I measured the possible social benefits of punishing in a PGG by using a simultaneous gift exchange instead of the Trust Game used in Study 2. This simultaneous gift exchange had an incentive structure similar to a Prisoner's Dilemma, but there was the option of sanctioning selfish partners. Also, instead of using the "strategy method" by having participants decide what monetary sums to trust to *each of the other three "players"*, I paired them with only one other "player" and asked them how much they wanted to send to *that particular player*. I also assessed whether participants remembered what the other "players" had done in the PGG.

4.5.1 Study 3 Methods

Seventy participants (52 women, average age 19.0 ± 1.3 years; 18 men, age 19.2 ± 1.2) played one round of the same PGG with punishment as in Study 2.

One of the computer players was a free-rider (contributed $1), one was a punisher (contributed $7, spent $2 to punish the free-rider), and one was a non-punisher (contributed $7, did not punish). The cost of sanctions was increased for this study (but only for this study), such that it cost $1 to reduce another person's payoff by $2.

After one round of the PGG with punishment, participants played a modified simultaneous gift exchange (also known as a "give-some" game; van Lange, Ouwerkerk, & Tazelaar, 2002). The "strategy method" was *not* used in this game to elicit decisions, so participants were paired with only one other participant (which was actually one of the computer players); 23 participants were paired with the free-rider, 24 were paired with the non-punisher, and 23 were paired with the punisher. Thus, Study 3 (unlike Study 2) was a Between-Subjects design. Each member of the "pair" was given $10, and they were simultaneously given the option of sending any number of these dollars to their respective partners; any amounts sent were doubled by the experimenter. Thus, the gift exchange was like a Prisoner's Dilemma, only with the option of sending any amount between $0 and $10 instead of having a binary Cooperate/Defect decision. After finding out the other "player's" decision, participants had the option to punish their partners. The reason for this second punishment stage was in order to give participants an incentive to send more to "players" who had already shown an unwillingness to tolerate free-riding (i.e. the punishers) in the punishment stage of the PGG. This was designed to test whether participants

would send more to punishers regardless of whether they did so to reward the punishers or to avoid being punished. After completing the simultaneous gift exchange and second punishment stage, participants' memory of the PGG was tested. Participants were asked what each of the other "players" had contributed and spent on sanctions (and whom they had sanctioned) in the PGG.

4.5.2 Study 3 Results

Participants contributed an average of \$6.02 (S.D = 2.82) in the PGG. Of the 70 participants, 30 (43%) punished one of the other "players" (almost always the free-rider), and the average amount spent on sanctions was \$0.71 (S.D. = 1.10). Punishment cost more in this study than in Study 2, so it is not surprising that there was slightly less punishment in Study 3. Men did not contribute more in the PGG than women did (\$5.7 vs. \$6.1, $t < 1$), but they spent more on sanctions than women did (\$1.3 vs. \$0.5, $t = 2.20$, $p = 0.039$). There was a positive correlation between how much a participant contributed in the PGG and how much he/she gave the other "player" in the simultaneous gift exchange ($r_{68} = 0.56$, $p < 0.001$). However, the amount that a participant spent on punishment was not significantly correlated with his/her PGG contributions ($r_{68} = 0.08$, n.s.) or the amounts he/she gave in the simultaneous gift exchange ($r_{68} = 0.14$, n.s.).

The effects in the simultaneous gift exchange were not as strong as the effects in the Trust Game in Study 2 (Figure 4.3). When participant sex was included in the analysis, free-riders were sent less than punishers and non-punishers combined ($F_{1,66} = 4.06$, $p = 0.048$), but there were no differences in the

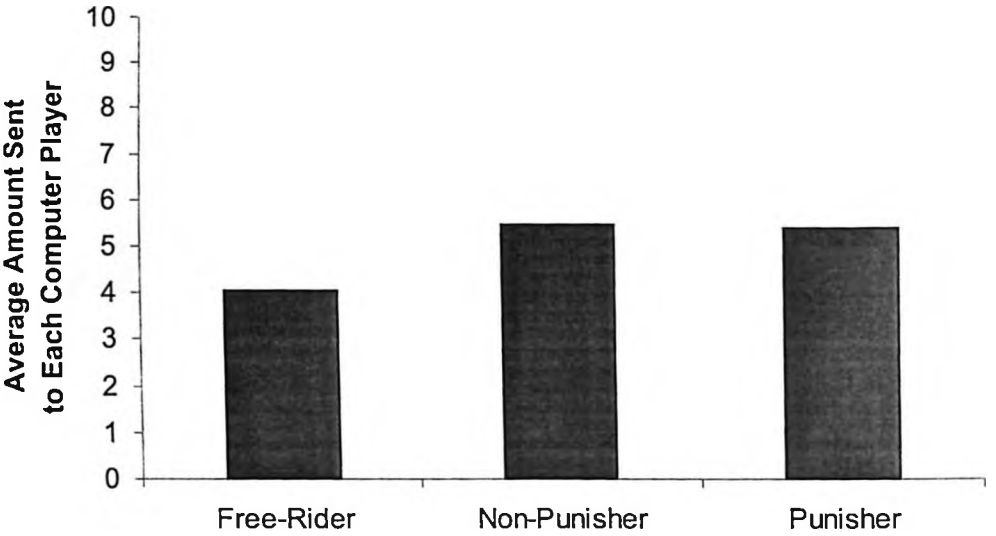


Figure 4.3: Average amounts sent to the free-riders, non-punishers, and punishers in the simultaneous gift exchange after one round of a public goods game in Study 3. Free-riders received less than punishers and non-punishers, yet there were no differences between punishers and non-punishers.

amounts sent to punishers and non-punishers ($F < 1$), as in Study 2. Participant sex did not have a main effect on amounts sent or an interaction with partner type (Fs < 1). Including participants' contributions in the PGG or punishment as covariates in the analysis did not change the results.

Data on memory for contributions and punishment were available for 62 participants. Most (92%) participants correctly remembered the pattern of others' contributions in the PGG. Only 60% (37/62) correctly remembered the pattern of punishment, while 18% (11/62) did not remember any punishment, 5% (3/62) remembered punishment by two other "players", 10% (6/62) mixed up which

other "player" was the punisher and which was the non-punishing contributor, and 8% (5/62) remembered some other incorrect pattern of punishment. However, the results did not differ after excluding the data from participants who did not correctly remember the punishment.

4.5.3 Study 3 Discussion

Study 3 replicated the results of Study 2 in that free-riders received less than contributors, and there was no difference in the amounts received by punishers and non-punishers. The differences between free-riders and contributors were smaller in this study than in the previous, and this could be because of the possibility of receiving punishment after the simultaneous gift exchange. Participants may have been reluctant to send very low amounts in the simultaneous gift exchange, even to free-riders, out of fear of sanctions after the simultaneous gift exchange. Study 3 also replicated Study 2 in suggesting that punishers were not trusted or rewarded less than non-punishers. This effect was not due to participants failing to remember sanctions, because the same pattern was present among the participants who did correctly remember.

It is curious that participants in Study 1 rated punishers as more trustworthy, but Studies 2 and 3 failed to find that punishers were trusted or rewarded more than non-punishers. It is possible that because there was only one round, participants did not yet have expertise in the game or strong emotional responses to the actions of others, such as anger towards a repeated free-rider. One round does not give players enough information to determine whether a free-

rider mistakenly made a low contribution or whether he/she will continue to contribute very little (which is more deserving of punishment). Also, with only one round, there was no chance for participants to see that sanctions induced the free-riders to cooperate, so they may not have realized that punishment of free-riders is beneficial to the group, and thus would not have felt gratitude or trust towards punishers. It is also still possible that seeing the identical contributions of the punisher and the non-punisher in the PGG made participants feel they should treat them equally, even though the participants were only paired with one other "player" in the simultaneous gift exchange. Study 4 addressed these possibilities.

4.6 Study 4

Study 4 tested whether participants would preferentially trust punishers after repeated exposure to a free-rider. In order to test whether punishers would only receive benefits from other punishers, I also examined whether participants' own punitive behaviour was related to their tendencies to trust (or distrust) punishers. Five rounds of PGG experience were sufficient to elicit positive perceptions of punishers in Study 1, so participants in Study 4 played five rounds of a PGG in which there was a strong free-rider, a punisher, and a non-punisher, and then played the same Trust Game as in Study 2. The punisher and non-punisher computer players were designed to contribute the same amounts in the PGG but in slightly different orders (for realism), so only the sanctioning behaviour would differ between them. The free-rider's PGG contributions

increased towards the end to simulate the effects of being punished, and the punisher was designed to wait a round before sanctioning and to stop after the free-rider started contributing. These computer players were designed to imitate the behaviour of freely interacting people as seen in other chapters and previous studies of this chapter.

4.6.1 Study 4 Methods

Fourteen females (mean age: 18.9 years, S.D. = 1.4 years) and 13 males (mean age: 19.0 years, S.D. = 1.1 years) played five rounds of a PGG with punishment against what they thought were real participants, but were actually preprogrammed computer players. One computer "player" was a free-rider who contributed \$1, \$0, \$1, \$3, and \$6, respectively, in the five rounds. The other two computer players were relatively cooperative and contributed the same total amount in the PGG, but one (henceforth "the punisher") punished the free-rider by \$0, \$2, \$2, \$3, and \$1, respectively, in the five rounds, whereas the other (henceforth "the non-punisher") never punished. Half of the participants saw the punisher contribute \$7, \$8, \$7, \$6, and \$8 in the five rounds and the non-punisher contribute \$7, \$7, \$6, \$8, and \$8; the contributions of the punisher and the non-punisher were switched for the other participants. The computer "players" did not change their behaviour in response to participants' contributions or punishment Participants were not told the number of rounds.

After completing the PGG, participants played the same modified version of Berg et al.'s (1995) Trust Game described in Study 2. As in Study 2, I used the

strategy method to gather data on how much participants would trust each of the other players by asking participants how much money they would entrust to each potential recipient (presentation order was randomized), and these decisions were binding because participants were told that they would be randomly paired and the corresponding trust decision would be implemented if they were assigned to be the Truster in their pair. Thus, the amount entrusted to different "players" in the Trust Game was a within-subject factor that was analyzed with a Repeated-Measures General Linear Model (SPSS 11.0). I used a median split to categorize participants as high or low punishers based on the amounts that they spent on free-rider punishment, and this between-subjects categorical variable was added to the analysis of trust. After making trusting decisions, all participants were told that they were assigned to be Responders and that the Truster had sent them $8, and were asked how much of the tripled amount ($24) they wished to return. Because the "strategy method" was not used to measure Responder behaviour, there were far fewer data points for that variable; there were 11 Responders for the free-rider, 8 for the non-punisher, and 8 for the punisher, and this was a Between-Subjects factor.

4.6.2 Study 4 Results

4.6.2(1) Public Goods Game:

In the five rounds, participants contributed an average of $6.3, $5.5, $5.6, $5.6, and $5.8 (respectively). No participants contributed less than the free-rider in the first three rounds when the free-rider's contributions were lowest, and this

is an important requirement for the free-rider to be perceived by all participants as being a low contributor. There was no significant change in participants' contributions between rounds and no linear decrease ($F < 1$); these results are consistent with other studies that include punishment in public goods games. There was no sex difference in contributions ($F < 1$).

Across the five rounds, participants spent an average of \$0.6, \$0.9, \$1.2, \$0.7, and \$0.1 on punishment, and there was a significant linear ($F_{1,26} = 4.74, p = 0.039$) and quadratic ($F_{1,26} = 12.03, p = 0.002$) component to this pattern. This suggests that participants increased their punishment until the free-rider started contributing more, because punishment started decreasing in the same round that the free-rider started contributing more (round 4). Of the 27 participants, 20 punished at least once. Fifteen of those 20 punished the free-rider exclusively, 3 punished all three other "players" in one or two rounds, 1 mostly punished the free-rider but also punished the non-punisher in one round, and another exclusively punished the punisher in one round. Participants punished the free-rider more than the non-punisher and punisher (means: \$3.1 vs. \$0.3 and \$0.2, respectively, orthogonal contrast: $F_{1,26} = 26.43, p < 0.001$), and did not differentially punish the latter two ($F < 1$). Although men spent more on punishment than women (means: \$4.5 vs. \$2.7), this difference did not reach significance ($F_{1,25} = 2.32, p = 0.14$).

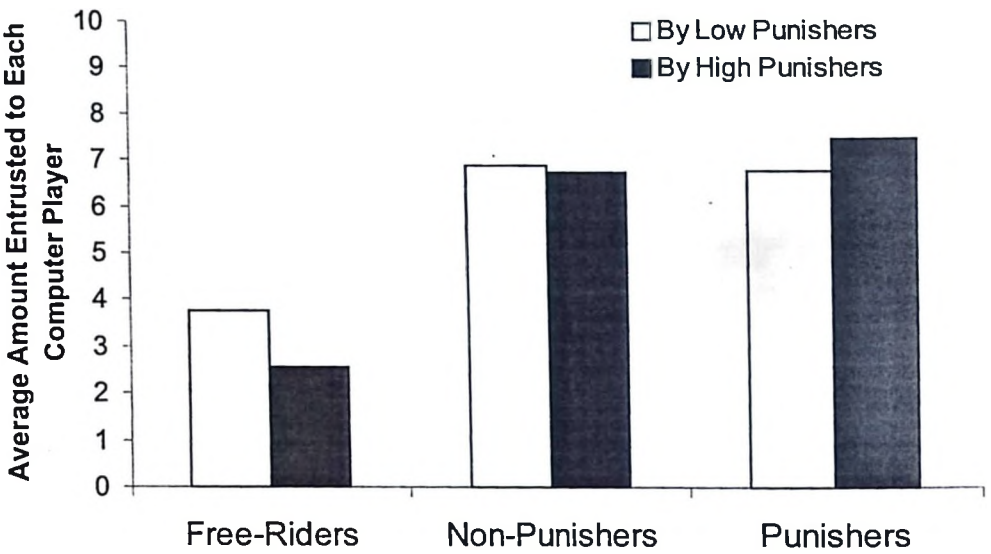


Figure 4.4: Average amounts entrusted to free-riders, non-punishers, and punishers in the Trust Game after five rounds of public goods game in Study 4, by participants who provided more (black bars) or less (white bars) than the median amount of punishment. Free-riders received less than cooperators, and punishers received more than non-punishers.

4.6.2(2) Trust Game:

There were significant differences between the amounts entrusted to free-riders, punishers, and non-punishers ($F_{2,50} = 43.36$, $p < 0.001$, see Figure 4.4). An analysis of orthogonal contrasts revealed that free-riders were trusted less than punishers and non-punishers ($F_{1,25} = 45.55$, $p < 0.001$), and punishers were trusted significantly more than non-punishers ($F_{1,25} = 4.34$, $p = 0.048$). Eleven participants entrusted different amounts to the punisher and non-punisher, and 9 of these 11 trusted the punisher more (binomial test: $p = 0.033$). There was no interaction between participants' own punishment and the amounts they entrusted

to all three other "players" ($F_{2,50} = 1.98, p = 0.15$). However, the orthogonal contrast was significant ($F_{1,25} = 6.69, p = 0.016$) for the interaction between participants' own punishment and the amounts entrusted to non-punishers versus punishers. An analysis of this interaction reveals that lower-than-average punishers did not entrust different amounts to punishers and non-punishers ($F < 1$), but higher-than-average punishers trusted punishers more than non-punishers ($F_{1,13} = 8.55, p = 0.012$). Including participant sex did not alter any of these results.

Participants' PGG contributions were significantly correlated with the amounts they entrusted to others in the Trust Game ($r_{25} = 0.57, p = 0.002$). The amounts that participants returned were significantly correlated with the amounts they themselves entrusted ($r_{25} = 0.58, p = 0.001$), but were not correlated with either their total contributions or punishment ($rs_{25} = 0.21$ and -0.28, both n.s.). Different amounts were returned to different "players" ($F_{2,24} = 3.58, p = 0.043$). Contrast analysis revealed that participants returned less money to free-riders than to non-punishers and punishers (\$8.1 vs. \$12.9 and \$10.3, $p = 0.03$), but did not return different amounts to punishers and non-punishers ($p = 0.18$). Adding a participant's average trusting behaviour as a covariate produced the same results.

4.6.3 Study 4 Discussion

This study replicated the other studies by showing that people trusted the free-rider less than the other "players". More importantly, it showed that punishers were trusted more on average than non-punishers after five rounds of a

PGG. Most participants trusted the punisher and non-punisher equally, but this is not surprising given that their total contributions were identical and contributions were an important determinant of trust. In fact, their contributions and punishment only differed in the middle rounds, so primacy and recency effects would make them seem similar. Of the participants who treated them differently, a significant number trusted the punisher more.

The structure of Study 4 was identical to that of Study 2, only with more rounds and thus longer interaction, so this is likely to have been the factor that caused participants to trust punishers in Studies 1 and 4, but not in Studies 2 or 3. The computerized free-rider in Study 4 was arguably more "deserving" of sanctions than in Studies 2 and 3, because it continued to contribute relatively little and only increased contributions in response to punishment, and many participants spontaneously commented on that fact after the experiment. Also, participants could observe the effects of sanctions on the free-rider's behaviour, and could note that the punisher did not punish anyone else. Study 4 did not determine which of these factors is the most important element of repeated interactions, but did show the effects of prolonged exposure to free-riders and punishers. It is interesting that punishers were trusted more, but were not returned more money in the Trust Game. This suggests that people may trust punishers, but do not reward them more than non-punishers. If so, this would support the idea that people treat punishment as a signal of trustworthiness (despite the fact that

participants' punishment was not correlated with the amounts they returned in this experiment).

Having multiple rounds allows for the possibility of "second-order punishing" (i.e. punishing non-punishers), which several theorists suggest is a significant force in maintaining the existence of punishment and cooperation (e.g. Bendor & Swistak, 2001; Henrich & Boyd, 2001; Sober & Wilson, 1998). Even though each participant in this study saw a clear free-rider that one "player" failed to punish, there was no evidence that participants engaged in second-order punishing. Participants punished the non-punisher as often as the punisher, and there was no difference in the amounts they were punished. This is consistent with the results of Kiyonari, Shimoma, and Yamagishi (2004) and Kiyonari and Barclay (2005), and seriously weakens any theoretical models that rely on second-order punishment to maintain the existence of punishment.

4.7 Study 5

Study 5 sought to replicate the findings of Study 4, and test whether they would occur in PGG games with naturally occurring variation. Participants in Study 5 played five rounds of a PGG and then the Trust Game from Studies 2 and 4. Participants experienced naturally occurring variation in cooperation and punishment because computer players were not used in this study. If the effects of punishment on trustworthiness are similar, then this naturally occurring variation allows us to generalize the results a bit more than we could with artificially

occurring variation alone, and it allows us to examine different types of punishment. Some natural punishment is arguably justifiable because it is directed at free-riders, while some is not because it is directed at cooperators. In Study 5 it was possible to examine the differential effects of justified and unjustified punishment on people's trustworthiness.

4.7.1 Study 5 Methods

4.7.1(1) Participants and Procedure

Fifteen males (average age 20.6 ± 4.5 years) and forty-five females (average age 18.8 ± 1.0 years) played five rounds of a PGG with punishment in 15 groups (of four participants each) without computer players, and then played the modified version of Berg et al.'s (1995) Trust Game described in Studies 2 and 4. As in Studies 2 and 4, I gathered data on how much participants would trust each of the other players by asking participants how much money they would entrust to each potential recipient (strategy method), and I randomly formed the pairs afterwards. In each pair, the participant who was assigned to be the Responder was told what he/she received from the Truster, and then decided how much of the tripled amount to return to the Truster. Thus, the strategy method was not used for Responder decisions, so there were far fewer data points for this variable (30 total, each with a different amount sent).

4.7.1(2) Statistical Analysis:

Punishment was coded as either justified or unjustified according to a defined algorithm. Justified punishment was defined as sanctions imposed on the

lowest contributor (the "free-rider") in the group on any given round, or punishment from the bottom up if more than one person was punished, as long as the punished person had contributed less than the punisher. Unjustified punishment was defined as any other sanctions, which included retaliation, punishment of people who contributed as much as or more than the punisher in the given round, and punishment of a low contributor if there was someone else who contributed less but was not punished.

To test whether the participants trusted justified punishers more than players who did not provide justified punishment, I conducted Multiple Linear Regression analyses (SPSS version 12.0) to see what factors would predict the amount each participant was entrusted with. For each recipient, I examined the average amount of money that the other group members were willing to entrust to him/her. Past research has found that there is a correlation between the total PGG contributions in a group and the level of trust displayed by group members in subsequent Trust Games (Barclay, 2004; Chapter 2). Thus, it is important to control for group trusting behaviour, because groups with strong free-riders are likely to require justified punishment, but the low PGG contributions by the free-rider will also bring down the group's level of trust. For this reason, I factored out the general levels of trust exhibited by other players by creating dummy variables for each group. Such dummy variables factor out the general levels of trust exhibited by each group in order to compare each person to his/her group, which is the relevant comparison group to test against when testing for a within-group advantage or disadvantage of punishing. The regression model for predicting the average amount (A) entrusted to any individual i in group j was thus:

$$A_{ij} = b_0 + b_1c_i + b_2J_i + b_3U_i + \Sigma_{(j\,=\,1\ \text{to}\ 15)}\, b_j g_{ji} \qquad (1)$$

where b_0, b_1, b_2, and b_3 are constants, c_i is the individual's level of PGG contributions, J_i is his/her justified punishment, U_i is his/her unjustified punishment, and g is the dummy variable for group membership such that $g_{ji} = 1$ if the individual is in group j and 0 otherwise (i.e. it is the group-specific "effect").

4.7.2 Study 5 Results

4.7.2(1) Public Goods Game:

In the five rounds, participants contributed an average of \$5.5, \$6.0, \$7.0, \$7.5, and \$7.8 (respectively), which represents a significant linear increase (orthogonal contrast: $F_{1, 14} = 25.84$, $p < 0.001$). Of the 60 participants, 23 provided no punishment, 19 provided only justified punishment (average \$2.9 spent), 7 provided only unjustified punishment (average \$3.6 spent), and 11 provided both (average \$3.3 and \$3.4 spent, respectively). There were 64 instances of justified punishment being delivered against a participant, and 49 instances of unjustified punishment. Of the latter instances, 23 were retaliation for punishment in the previous round[9], 6 were retaliation delayed by one round, 12 were free-riders punishing everyone (9 of which occurred while retaliating against someone else), 3 were free-riders punishing the highest contributor alone[10], 2 were free-riders punishing the second-lowest contributors ("hypocritical punishment"), 2 were delayed punishment of free-riders, and 1 was a participant

[9] Some of this was retaliation by non-free-riders against unjustified punishment, which is arguably justifiable, but it makes no difference whether these are coded as justified or unjustified.

[10] Punishment of contributors is often found at non-zero levels (e.g. Fehr & Gächter, 2000), and has been interpreted as generalized or preemptive retaliation.

apparently retaliating on behalf of someone else. There were no unambiguous cases of people punishing non-punishers (second-order punishment). Punishment did not change in frequency across rounds ($F < 1$).

Men did not contribute more in the PGG than women did ($36.1 vs. $32.9, $t = 1.10$, $p = 0.28$), but they spent more on sanctions than did women ($4.4 vs. $1.9, $t = 2.72$, $p = 0.009$). This was because men spent more on justified punishment than women ($2.7 vs. $1.1, $t = 2.23$, $p = 0.026$), although not significantly more on unjustified punishment ($1.7 vs. $0.8, $t = 1.27$, $p = 0.22$). After controlling for group contributions, a participant's PGG contributions were positively correlated with the amount of justified punishment he/she provided ($r = 0.33$, $p = 0.011$) and negatively correlated with unjustified punishment ($r = -0.36$, $p = 0.006$).

4.7.2(2) Trust Game:

Group trusting behaviour was an important determinant of amounts received in the Trust Game, because all of the dummy variables for group membership were significant (average Standardized β for dummy variables = 0.548, all ps < 0.05). Justified punishment positively predicted the amounts received in the Trust Game (Standardized $\beta = 0.232$, $t = 2.21$, $p = 0.032$), whereas unjustified punishment negatively predicted the amounts received (Standardized $\beta = -0.227$, $t = -2.23$ $p = 0.031$) and contributions in the PGG were not a significant predictor (Standardized $\beta = -0.019$, $t = -0.130$, $p = 0.90$). The regression model was highly significant ($F_{17,42} = 12.54$, $p < 0.001$) and accounted for 77% of the

variance in amounts received. Amounts sent to people in the Trust Game (A) can be described by the estimated equation:

$$A = 2.644 - 0.004\ c + 0.196\ J - 0.210\ U + G + \varepsilon \qquad (2)$$

where c is the recipient's total contributions, J is his/her justified punishment, U is his/her unjustified punishment, G is the group-specific "effect" (i.e. the group dummy variable and its coefficient; these have an average value of 4.199), and ε is unaccounted variance (error).

One potential problem with these analyses is that free-riders in a group cannot be justified punishers because the definition of justified punishment precluded participants who punished anyone that contributed more than themselves. Therefore, justified punishers may be entrusted with more money than non-punishers simply because participants did not entrust money to free-riders, and justified punishers could not be free-riders. However, recipients' justified punishment was a stronger predictor of amounts received than PGG contributions were, which speaks against this potential criticism. To further examine this potential problem, I ran the analysis again after excluding the lowest contributor in each group, who by definition could not be a justified punisher, leaving 45/60 participants. The regression model was still significant ($F_{17,27} = 8.35$, $p < 0.001$) and accounted for 74% of the variance in amounts received. Even with this reduced power, justified punishment predicted amounts received, although this just failed to reach significance (Standardized $\beta = 0.319$, $t = 1.98$, $p = 0.058$). The fact that it was a marginally significant positive predictor of

amounts received (even after reducing power by excluding the lowest contributor in each group) suggests that justified punishers were not trusted more merely because they were not free-riders. Neither unjustified punishment (Standardized $\beta = -0.069$, $t = -0.54$, $p = 0.59$) nor PGG contributions (Standardized $\beta = -0.154$, $t = -0.65$, $p = 0.52$) predicted amounts received in this reduced sample. The amounts sent to non-free-riders in the Trust Game (A) can be described by the estimated equation

$$A = 5.315 - 0.033\,c + 0.312\,J - 0.084\,U + G + \varepsilon \qquad (3)$$

Equation 3 uses the same symbols as Equations 1 and 2, except that the average group-specific coefficient (g) has a value of 2.232.

The amount sent by a Truster significantly predicted the amount returned to him/her in the Trust Game (Standardized $\beta = 0.447$, $p = 0.033$). Although the regression model was significant ($F_{18,11} = 2.69$, $p = 0.049$) and accounted for 51.2% of the variance in amounts returned to Trusters, no other variables significantly predicted amounts returned to Trusters, probably because there were far fewer observations to test so many variables (including group dummy variables); only half of the participants were Responders, they only returned money to one Truster each, and two Trusters did not send anything so there are no data for them or their Responders. The only thing that significantly predicted how much money a given Responder returned was the amount that the Truster sent to him/her (Standardized $\beta = 0.615$, $p = 0.002$), and this model was significant ($F_{18,11} = 3.82$, $p = 0.012$) and accounted for 63.7% of the variance in Responder

behaviour. Responders returned an average of 46% of the tripled amount that they received.

4.7.3 Study 5 Discussion

This study replicated some results from previous studies. Previous studies have shown that opportunities to punish others and gain a reputation for trustworthiness can cause PGG contributions to increase across rounds (e.g. Barclay, 2004, Chapter 2; Fehr & Gächter, 2002), and this study showed that the two effects together cause an increase in contributions. This study also found that men punished more than women. This was not simply because men were more aggressive than women, as men provided more justified punishment but not more unjustified punishment. Fehr and Gächter have suggested that some punishment of cooperators is retaliatory punishment, and this study shows that retaliatory punishment occurs frequently when participants can see who imposed sanctions. Approximately 20% (23/113) of the instances of punishment were apparently in retaliation for some form of punishment. The frequency of this retaliatory punishment suggests that punishment is not costless, as some theorists have argued (e.g. Sober & Wilson, 1998).

This study also replicates Study 4 in failing to find unambiguous evidence for second-order punishing (punishing non-punishers). Theorists have predicted that this is an important component of the evolution of cooperation (e.g. Henrich & Boyd, 2001; Sober & Wilson, 1998), but other studies have also noted a

conspicuous lack of second-order punishment (Kiyonari & Barclay, 2005; Kiyonari, Shimoma, & Yamagishi, 2004). If second-order punishment is not common, then the evolution of punishment must have been supported by other processes, such as increased trustworthiness of punishers.

After controlling for the trusting behaviour of other group members, justified punishment did significantly predict how much a person was trusted with. Justified punishers were even trusted somewhat more than other cooperators, which suggests that the effect was not simply because justified punishers were trusted more than free-riders. Thus, this study replicates Study 4 by showing that people trust more money to those who have demonstrated justified punishment, although neither study provided evidence that people return more money to them. Not all punishment led to trustworthiness, because recipients' unjustified punishment negatively predicted how much they were entrusted with. It is somewhat surprising that participant's PGG contributions did not predict the amounts they were trusted with, because this was found consistently in Studies 2-4 and in past research (Barclay, 2004, Chapter 2). People who made high PGG contributions tended to provide more justified and less unjustified punishment, and both of those were significant predictors of amounts received in the Trust Game. I suspect that those two variables together accounted for the trust that would otherwise have been attributed to PGG contributions, such that contributions did not predict anything beyond that which was predicted by punishment. However, contributions are probably a necessary component of the

trustworthiness signal, such that punishment without contributions is seen as hypocrisy and is judged unjustified.

This study cannot speak to the question of whether justified punishers actually were more trustworthy than non-punishers because there were relatively few data points on Responder behaviour. Justifed punishers may be very discriminating in their trustworthiness, such that they repay the trust of cooperators but not of free-riders, just as high contributors tend to do (Albert, Güth, Kirchler, & Maciejovsky, 2002). This is especially likely to happen with punishers, because the act of punishment demonstrates a dislike of free-riders that could easily cause them to repay the trust of free-riders less than non-punishers would. Thus, punishers might not be more trustworthy overall, but only towards cooperators, and future studies could investigate this by varying the level of cooperativeness of punisher's partners.

4.8 General Discussion

Study 1 demonstrated that people rated altruistic punishers as more trustworthy, group-focused and worthy of respect than non-punishers. The participants in that experiment were familiar with the situation because they had just experienced a free rider in a public goods game. Studies 4 and 5 supported this by finding that justified punishers were trusted more than non-punishers and received monetary benefits for punishing. However, Studies 2 and 3 did not find

such an effect. Thus, we must look at the differences between these studies to find out what was likely to have caused the differences in results.

Studies 2, 3, and 4 all had computer players, so the naturally occurring variation in Study 5 could not be the only cause of differences. In Studies 2 and 3, seeing a punisher and a non-punisher both contribute 7 of 10 dollars may have made participants feel that they should treat them equally. Study 5 used naturally occurring variation in contributions and punishment, and so would not have had this potential problem. However, this is not likely to be the crucial difference for two reasons. First, punishers had the same total contributions as non-punishers in Study 2 and in Study 4, yet participants in Study 2 were split 4/3 as to whom they trusted more, whereas participants in Study 4 were split 9/2. Secondly, there was not even a hint that punishers were given more money than non-punishers in Study 3 even though participants never had to decide what to give to both, and thus would have no reason to consciously compare them.

The most likely explanation for the different results is that participants played five rounds of a PGG in Studies 1, 4, and 5, but only one round in Studies 2 and 3. Punishment increased perceived trustworthiness in all the studies where there were repeated interactions (1, 4, & 5), but did not in any of the studies with only one round (2, 3, and the one-round PGGs in Barclay & Kiyonari, 2005, and Kiyonari, Shimoma, & Yamagishi, 2004). Five rounds allow for more time to gain expertise in the game and to experience emotional responses towards free-riders and punishers. After only one round, it may be too early to guess the

motivations of the free-riders and the punishers and to tell whether punishment of a free-rider is truly justified. In Study 5, unjustified punishment had a negative effect on one's reputation, and justified punishment had a positive effect, so if participants were unclear about the justification for the punishment in Studies 2 and 3, then these positive and negative effects would cancel each other out. Also, five rounds allow participants to appreciate the effects of sanctions on the behaviour of the free-riders, which may be necessary for people to trust punishers. For these reasons, the studies with five rounds may provide a better test of the effects of punishment on one's reputation. Future studies can be done to see whether punishment needs to be effective in order to bring reputational benefits.

Together, these results suggest that costly sanctioning of free-riders might not actually be costly once there are opportunities for the punisher to acquire a reputation. Punishers provide a public good by forcing free-riders to cooperate, and people do seem to realize this after playing multiple rounds of a PGG. Whether this makes up for the costs of punishment depends on the frequency of collective action projects (and free-riders to punish) and dyadic opportunities for trust outside the laboratory. If dyadic interactions are more frequent or carry larger potential payoffs than collective action projects, then the reputational benefits of punishing could easily compensate the punishers for more than the cost of the altruistic punishment, such that justified punishers actually do better than non-punishers. Future studies should test whether punishers are similarly rewarded or trusted outside the laboratory. If so, then reputation could eliminate

the disincentive to punish free-riders, and cause punishment to increase in frequency in populations via individual learning. If such benefits were also accrued in ancestral environments in which humans evolved, then reputation (with or without group-level effects) could explain why the psychological mechanisms that modulate altruism and altruistic punishment evolved.

Chapter 5 Altruism as a courtship display: Effects of altruism on audience perceptions

Abstract

Costly signaling theory suggests that altruistic individuals are expected to be desirable as romantic partners. Some studies have found that altruists are less desirable than "heroes". Other studies showed that altruism is attractive by contrasting descriptions of "nice guys" with "jerks". The present experiments sought to resolve this debate by having participants read fictive vignettes of persons with corresponding photographs, such that altruistic vignettes were compared with control descriptions that differed only in the presence or absence of small hints of altruistic tendencies. In one experiment, women rated self-reportedly altruistic men as more desirable partners than matched controls, whereas men rated altruistic women as being less attractive than matched controls. In a similar experiment with descriptions by third parties, women preferred altruistic men for certain relationship categories, but men exhibited no preference. These results are discussed with regard to the idea that altruism by males may serve as a courtship display that honestly signals good character.

5.1 Introduction

Altruistic behaviour is behaviour that benefits others at a cost to oneself. Since Darwin, the existence of altruism has been an apparent problem for evolutionary theory. Kin selection (Hamilton, 1964) and reciprocal altruism (Alexander, 1987; Trivers, 1971) have gone a long way in explaining how altruism can evolve in a world of selfish genes. In humans however, there are many instances of altruism that cannot be explained by either kin selection or reciprocal altruism because they involve generosity towards non-kin where the altruist cannot target his or her generosity specifically towards reciprocators. Recently, costly signaling theory (Zahavi & Zahavi, 1997) has been used to explain altruistic behaviours such as provisioning for feasts (e.g. Boone, 1997; Smith & Bliege Bird, 2000). By giving benefits to others, an altruist is proving that he or she is of high enough quality or status to bear the costs of conferring those benefits. Tessman (1995) suggested that human altruism is a courtship display that honestly signals an individual's ability and willingness to be a good parent, and Miller (2000) argued that altruism can act like a peacock's tail as a costly display of abilities and resources. Altruism towards a potential mate is likely to be attractive because it signals interest and concern for that potential mate, but altruism towards third parties can also be attractive if it signals abilities, resources, or good character. These theories all predict that people should be able to detect altruism, and should find altruists more attractive than non-altruists. If altruism signals abilities, and such abilities are considered attractive (see for

example Faurie, Pontier, & Raymond, 2004) then it should affect perceived attractiveness and desirability for short- and long-term relationships. Altruism that signals character (but not abilities) should be an attractive quality in long-term partners, but not necessarily in short-term partners because of the lack of opportunity to benefit from a partner's good character, and so it is less likely to affect perceived attractiveness.

One might expect that women would be more sensitive to cues of altruism than men for a few reasons. Women are often more choosy in their mating partners (Gaulin & McBurney, 2001), and given that, men are more likely than women to signal abilities to attract mates (Miller, 2000). Although both sexes are concerned with good character in many cultures (Buss et al., 1999), women should be more concerned with the good character of men in order to avoid abandonment or violent relationships, because these are more problematic for females than males (Alcock, 1993; Daly & Wilson, 1988). For these reasons, women may be more attracted than men to altruistic tendencies in mates.

Folk wisdom apparently argues otherwise, suggesting that "nice guys" are less attractive than "bad boys". However, this is often an unfair comparison because "nice guys" and "bad boys" may differ on many dimensions other than niceness. Jensen-Campbell, Graziano, and West (1995) presented videos of males that varied in agreeableness and dominance, and found that women rated agreeable or prosocial men as more attractive than disagreeable men, especially when the men acted dominantly instead of subordinately. Similarly, Mims,

Hartnett and Nay (1975) found that men were rated more positively after being observed acting nicely than after acting obnoxiously. However, these studies do not show that women are attracted to men who behave altruistically towards other people (third-party altruism), nor do they conclusively show that a "nice guy" is more desirable than a neutral guy. Instead, they show that "nice guys" are preferred to "jerks". In order to show that third-party altruism is desirable, a study needs to provide a neutral condition to see whether altruism can raise an individual's desirability. The absence of a neutral control when contrasting "nice guys" with "jerks" confounds the interpretation of women valuing niceness or altruism with women disliking jerks. Women might actually prefer neutral men to either "nice guys" or "jerks", just as Burger & Cosby (1999) found that both dominant and submissive men are less attractive than men who display neither trait. In fact, Urbaniak and Kilmann (2003) found that women rated a particular man as being more desirable when he was portrayed as being "nice" rather than a "jerk", but there was little difference between the "nice" and the "neutral" guy. However, they only used one example of each so the effects may not generalize to other instances or other men.

The use of single examples was also a problem in a study by Kelly and Dunbar (2001). They found that heroism was a more important factor than altruism in women's mate choice, but altruism did seem to have some impact, such that altruists were more desirable than non-altruists for long-term relationships but not short-term relationships. However, Kelly and Dunbar varied

three factors (altruism, courage, and professional/volunteer engagement in such acts), and used one fictive description for each combination of those factors. For example, the sole fictive description describing a man whose job involved risky altruistic acts was completely different from that of the man whose job involved risky non-altruistic acts. We cannot generalize such results when they come from comparisons of single descriptions that differ in factors other than the supposedly focal factors.

The present study examined men's and women's attraction to opposite-sex photographs that were accompanied by descriptions that varied in the level of altruism described. Ratings of attractiveness for altruists were compared to the attractiveness of the same ads without mention of altruism. I did this for four different advertisements so that the findings would be more generalizable than in previous studies. I also varied the level of commitment sought by the people in the descriptions (the target), because people may prefer different traits in short-term partners than long-term partners (e.g. Gangestad & Simpson, 2000). Experiment One measured attraction to self-reported altruism using simulated dating advertisements. The context of a dating service was used so that it would put participants in a mate selection mindset.

5.2 Experiment One

5.2.1 Experiment One Methods

5.2.1(2) Participants. Seventy-five female (mean age = 20.1 ± S.D. 2.1 years) and seventy-five male (mean age = 19.7 ± S.D. 2.0 years) undergraduates

were recruited from an undergraduate psychology course at McMaster University as part of their course requirements.

5.2.1(2) Stimuli and Procedure. Simulated dating advertisements were created from phrases used in actual online dating services. Each simulated advertisement had a control version, and an "altruistic version" that differed only in the addition of a short descriptive phrase implying altruistic tendencies (e.g. "... and I enjoy helping people") and a hobby that also implied altruism (e.g. volunteering at a food bank). Based on a pilot study, four of these ads were selected, with two ads "seeking" short-term relationships and two ads "seeking" long-term relationships (see electronic supplementary material), each with an altruistic and control version. Pictures were downloaded in 2001 from an Internet site where pictures are rated for attractiveness (www.amihotornot.com). Upper body photographs of university-age men and women were selected if they had attractiveness ratings equal to the median for their sex (men: 8.3; women: 8.0; based on 60 pictures each). Four pictures for each sex were used, and these were counterbalanced across the four ads (and two versions of each ad), with the order of presentation randomized.

Each participant received a package that contained all four ads: the two short-term-seeking ads (one in the altruistic version and the other in the non-altruistic control version) and the two long-term-seeking ads (also with one altruist and one non-altruist). Thus, each participant saw a non-altruist and an altruist each seeking a short-term relationship, and a non-altruist and an altruist

each seeking a long-term relationship. The same ads were used for both sexes, but each participant only saw pictures of the opposite sex. Participants rated each picture with respect to their willingness to date, to have a long-term relationship, to work with the target, to be a platonic friend, or to lend money to the target. Participants then rated each target on physical and sexual attractiveness and many personality traits (used as fillers). All ratings were completed for each target before continuing to the next vignette. Questions were presented in five different random orders to reduce order effects. To minimize picture effects, scores on each dependent variable were standardized according to the mean and standard deviation for each picture on that dependent variable.

After rating all four targets, participants completed the Sociosexual Orientation Inventory (SOI; Simpson & Gangestad, 1991). This questionnaire assesses where participants fall on a continuum between a stated willingness to engage in (and approval of) "unrestricted" mating strategies characterized by multiple sexual relationships with low commitment (high scores) or "restricted" mating strategies characterized by fewer sexual relationships and greater commitment (low scores). Within each sex, participants were divided into tertiles. The data were analyzed using a Repeated Measures General Linear Model (SPSS 12.0), with two within-subject variables (Altruism and Relationship Type of the target), and for the questions related to mating (i.e. long-term relationships and dates) SOI tertile was added as a between-subjects variable. If a participant did not answer a particular question, he/she was excluded from the analysis of that

variable. Five females and seven males did not complete the SOI, so they were excluded from the analysis of long-term relationships and dates.

5.2.2 Experiment One Results

5.2.2(1) Effects of Altruism on Desirability of Males. Figure 5.1 presents the effects of altruism on the desirability of targets, and Table 5.1 breaks down this information by Relationship Type. Female participants were more willing to have long-term relationships or dates with altruistic males than non-altruistic males ($Fs_{1,73}$ = 5.75 and 5.43, respectively, ps < .05). These remained significant when SOI was added to the analysis ($Fs_{1,66}$ = 5.65 and 5.11, respectively, ps < .05), and SOI did not interact with either of these variables (Fs < 1.5, n.s.). Women were also more likely to lend money to altruistic males ($F_{1,72}$ = 10.86, p < 0.005), enjoy working with them ($F_{1,73}$ = 8.18, p < 0.01), or have platonic friendships with them ($F_{1,73}$ = 13.53, p < 0.001). Altruistic males were not rated as being more physically attractive than non-altruistic males ($F_{1,74}$ = 1.48, n.s.), but they were rated as being more sexually attractive ($F_{1,73}$ = 4.84, p < 0.05). Target's Altruism and Relationship Type did not have an interactive effect on any variable (all n.s.) except on women's likelihood of enjoying working with the target ($F_{1,73}$ = 4.28, p < 0.05), such that women significantly preferred working with altruistic men over non-altruistic men if the man sought a short-term relationship ($F_{1,73}$ = 8.97, p < 0.005) but not if the man sought a long-term relationship (F < 1, n.s.).

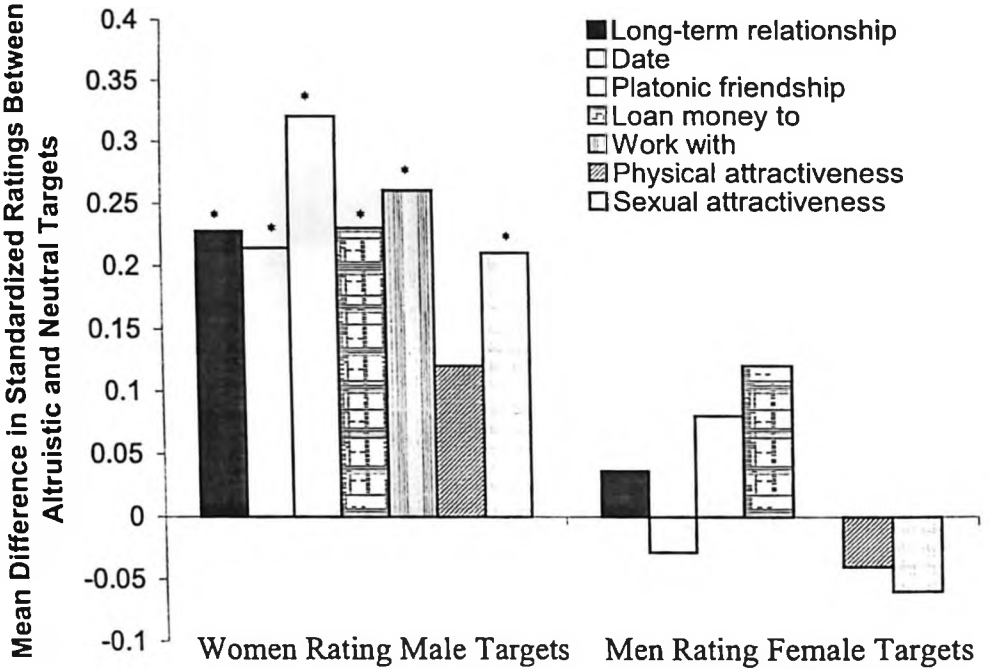


Figure 5.1: The effects of self-reported altruism (Experiment One) on the desirability of male and female targets for long term relationships (black bars), dates (grey bars), platonic friendships (white bars), loans (horizontally striped bars), working partnerships (vertically striped bars), and the effects of altruism on physical attractiveness (diagonally striped bars) and sexual attractiveness (dotted bars). Ratings were standardized according to the mean and standard deviation of each picture on each variable. * $p < 0.05$

Table 5.1: Participants' standardized mean ratings (and standard errors) of target men and women on the target's desirability and attractiveness in Experiment One. Ratings represent standard deviations from the sex-specific mean for each photograph on each variable.

	Target Seeking Short-term		Target Seeking Long-term	
	Neutral	Altruist	Neutral	Altruist
Long-Term Relationship				
Male Target	-0.42 (0.09)	-0.18 (0.11)	0.20 (0.12)	0.44 (0.12)
FemaleTarget	-0.26 (0.11)	-0.18 (0.12)	0.26 (0.12)	0.28 (0.13)
Date				
Male Target	-0.32 (0.13)	-0.02 (0.12)	0.11 (0.11)	0.26 (0.11)
FemaleTarget	0.00 (0.12)	-0.07 (0.13)	0.14 (0.12)	0.14 (0.11)
Platonic Friendship				
Male Target	-0.45 (0.12)	-0.06 (0.11)	0.12 (0.11)	0.39 (0.09)
FemaleTarget	-0.13 (0.11)	-0.03 (0.12)	0.06 (0.12)	0.09 (0.12)
Work Partnership				
Male Target	-0.49 (0.12)	-0.03 (0.12)	0.23 (0.10)	0.29 (0.10)
FemaleTarget	-0.14 (0.11)	-0.10 (0.12)	0.15 (0.11)	0.09 (0.11)
Loan Money				
Male Target	-0.48 (0.10)	-0.13 (0.12)	0.25 (0.10)	0.38 (0.12)
FemaleTarget	0.35 (0.10)	-0.21 (0.10)	0.23 (0.12)	0.33 (0.12)
Physical Attractiveness				
Male Target	0.06 (0.10)	0.08 (0.12)	-0.19 (0.11)	0.05 (0.12)
Female Target	0.01 (0.12)	0.04 (0.11)	0.03 (0.11)	-0.08 (0.12)
Sexual Attractiveness				
Male Target	-0.08 (0.10)	0.10 (0.12)	-0.12 (0.11)	0.11 (0.13)
FemaleTarget	0.07 (0.13)	-0.01 (0.11)	-0.01 (0.11)	-0.05 (0.12)

5.2.2(2) Effects of Altruism on Desirability of Females. There was no significant effect of altruism on men's willingness to have long-term relationships or dates with the target females, whether or not SOI was in the analysis (all Fs < 1, n.s., see Figure 5.1 and Table 5.1), and SOI did not interact with altruism in its effects on these variables (Fs < 1.6, n.s.). There was no effect of altruism on men's likelihood of lending money to the target females, enjoying working with them, or having platonic friendships with them (Fs$_{1,74}$ = 2.18, 0.00, and 0.55, respectively, all n.s.). Furthermore, altruism had no effect on the physical or sexual attractiveness of the female targets (both Fs < 1).

5.2.2(3) Effects of Relationship Type Sought. Female participants rated long-term-seeking targets more favourably than short-term-seeking targets with respect to their willingness to date, have long term relationships with, loan money to, enjoy working with, or have platonic friendships with the targets (all Fs > 9.0, all ps < 0.005). Male participants rated long-term-seeking targets more favourably than short-term-seeking targets with respect to their willingness to have long term relationships, work with, or loan money to the targets (all Fs > 4.5, all p < 0.05), but not with respect to their willingness to date or have platonic friendships with the targets (both Fs < 3.5, both n.s.). Targets seeking long-term relationships were not rated more physically or sexually attractive than targets seeking short-term relationships by either female participants ($F_{1,74}$ = 1.72 and $F_{1,73}$ = 0.01, respectively, both n.s) or male participants (both Fs < 1). One interpretation of these findings is that female participants felt that men who explicitly advertised

that they sought short-term relationships were somewhat sleazy, and male participants felt the same. There were many differences in the descriptions of the targets seeking long-term and short-term relationships, so these results might not generalize to all individuals who explicitly seek long-term or short-term relationships.

5.2.3 Experiment One Discussion

These results show that altruism increased men's attractiveness and dating desirability. Even a minor cue of altruism was sufficient to cause significant differences in sexual attractiveness and desirability for dates, long-term relationships, or platonic friendships. This supports the idea that male altruism can serve as a courtship display. By signaling ability and/or willingness to confer benefits upon others, a man can demonstrate his mate value, which could lead to higher reproductive success than a similar male who didn't signal. Claiming to regularly perform altruistic activities is not mere "cheap talk" if potential mates can verify such statements, as would have been possible with the types of altruism in this study (e.g. volunteering at food banks or Big Brothers/Big Sisters). This study did not distinguish among signals of ability, of mate quality (Smith & Bliege Bird, 2000), or of willingness to be a good partner for marriage, parenthood or work (Sosis, 2000; Tessman, 1995), but the results suggest that multiple signals may be occurring. Males' altruism affected not only women's willingness to have long-term relationships with them, but also males' sexual

attractiveness and women's willingness to have a single date with them, suggesting that the effect is not confined to long relationships.

If altruism is attractive to women because it signals mate quality, then it is plausible that altruism would be attractive to men also, because human males do seek evidence of quality more so than males of other species. Thus, it is interesting that altruism had no effect on men's preferences. This was probably not caused by men failing to pay attention to the stimuli, because men did pay enough attention to prefer to have long term relationships or work partnerships with long-term-seeking women. The manipulation of altruism was subtler than the manipulation of relationship type, so perhaps the effects of the latter overshadowed the effects of altruism for male participants. Another possibility is that men perceive self-reported female altruists as being "good girls" who would require much time and commitment in a relationship. This is unattractive to men who do not want much commitment in a relationship (Buss & Schmitt, 1993), and could counteract any positive effects of generosity in a woman's character. However, participants with high SOI scores did not react differently to altruism than participants with low SOI scores, which speaks against this hypothesis.

If announcing one's generosity has negative effects on one's desirability that counteract the positive effects of that generosity, then this might not occur with altruism that is not self-advertised. In order to attract a similar partner, one would presumably only advertise qualities in a dating advertisement if they were very important to oneself. Experiment Two examined the effects of altruism on

attractiveness when the acts are mentioned by a person other than the altruist, and sought to replicate and generalize the findings of Experiment One. Experiment Two used subtler manipulations of the type of relationship sought by the people in the descriptions in case there were qualitative differences between the targets seeking long-term versus short-term relationships which could have caused participants to prefer the former. Also, Experiment Two used manipulations of altruism that controlled for any differences in ability levels that might have been inferred between altruistic and control descriptions.

5.3 Experiment Two

5.3.1 Experiment Two Methods

5.3.1(1) Participants Eighty men (mean age = 19.18 ± 1.02) and eighty women (mean age = 19.15 ± 1.29) were recruited via posters from an undergraduate psychology course and participated as part of their course requirements.

5.3.1(2) Stimuli and procedure Photographs of four new median-attractiveness men and women of university age were chosen from the internet site (www.amihotornot.com); the pictures were of the targets' head and shoulders. I created simulated e-mail messages that contained third-party descriptions of people. Descriptions were designed so they could be used for either sex with appropriate changes in pronouns. Each description had an altruistic and a control (non-altruistic) version that differed in a small mention of altruistic behaviour. For

example, the control version of one description mentioned that the target played guitar in a local establishment, whereas the altruistic version said that the target played guitar at a children's hospital. Thus, I tried to equalize the skills and ability levels displayed in the altruistic and control versions. Each description also varied in the length of relationship sought by the target. Thus, there were four versions of each description: a non-altruist seeking a short term relationship, a non-altruist seeking long term, an altruist seeking short term, and an altruist seeking long term. Based on pilot testing, four descriptions (see electronic supplementary material) were chosen as having effective manipulations of altruism and relationship type sought.

Each participant saw all four descriptions (one of each version), each paired with one of the photographs of the opposite sex, with description/photograph pairings counterbalanced. Participants were asked to imagine that a friend had sent descriptions of potential blind dates. They then rated each "person" on a number of characteristics including desirability for relationships. Questions relating to good character and promiscuity were added (after the questions about desirability) to test the effectiveness of the manipulated features of the descriptions. To reduce picture effects, scores on each dependent variable were standardized according to the sex-specific mean and standard deviation for each picture on each dependent variable. After rating the targets, participants completed the SOI, and they were again divided into tertiles based on their SOI scores. A Repeated Measures General Linear Model was used to

analyze the standardized scores in the same manner as Experiment 1. If a participant did not answer any particular question, he/she was excluded from the analysis of that variable. Two females and seven males did not complete the SOI, so they were excluded from the analysis of long-term relationships and dates.

5.3.2 Experiment Two Results

5.3.2(1) Manipulation Check. Altruistic targets were rated significantly more "considerate" and "good" by women ($Fs_{1,78}$ = 12.40 and 10.86, respectively, both $ps < 0.005$, Table 5.2), and by men ($F_{1,78}$ = 14.75 and $F_{1,79}$ = 13.09, respectively, both $ps < 0.001$). Men and women both rated targets seeking short term relationships as being more "promiscuous" than targets seeking long term relationships ($Fs_{1,78}$ = 13.46 and 13.07 respectively, both $ps < 0.001$). Targets seeking short term relationships were thought to be more willing to engage in sexual intercourse without love or commitment than targets seeking long term relationships (women rating: $Fs_{1,79}$ = 24.55 and 33.63, respectively; men rating: $Fs_{1,79}$ = 19.05 and 9.56, respectively; all $ps < 0.005$). Thus, it appears that I successfully manipulated target altruism and relationship type sought.

5.3.2(2) Effects of Altruism on Desirability of Males. Figure 5.2 presents the effects of altruism on the desirability of targets, and Table 5.3 breaks down this information by Relationship Type. Female participants reported being more willing to have long-term relationships with altruistic males than non-altruistic

Table 5.2: Participants' standardized mean ratings (and standard errors) of target men and women on relevant personality characteristics (manipulation check) in Experiment Two. Ratings represent standard deviations from the sex-specific mean for each photograph on each variable.

	Target Seeking Short-term		Target Seeking Long-term	
	Neutral	Altruist	Neutral	Altruist
Considerate				
Male Target	-0.23 (0.11)	0.05 (0.12)	-0.12 (0.11)	0.34 (0.11)
FemaleTarget	-0.25 (0.11)	0.13 (0.12)	-0.13 (0.11)	0.25 (0.11)
Good				
Male Target	-0.26 (0.11)	0.03 (0.12)	-0.03 (0.10)	0.27 (0.11)
FemaleTarget	-0.24 (0.11)	0.05 (0.12)	-0.08 (0.10)	0.28 (0.11)
Promiscuous				
Male Target	0.32 (0.11)	0.07 (0.11)	-0.13 (0.10)	-0.28 (0.12)
FemaleTarget	0.09 (0.11)	0.29 (0.11)	-0.17 (0.10)	-0.21 (0.12)
Willingness to Have Sex Without Love				
Male Target	0.32 (0.11)	0.19 (0.11)	-0.17 (0.11)	-0.35 (0.11)
FemaleTarget	0.26 (0.11)	0.19 (0.11)	-0.15 (0.11)	-0.29 (0.11)
Closeness Required Before Sex				
Male Target	0.30 (0.12)	0.25 (0.11)	-0.21 (0.10)	-0.34 (0.10)
FemaleTarget	0.18 (0.12)	0.17 (0.11)	-0.17 (0.10)	-0.18 (0.10)

males. This effect just failed to reach significance on its own ($F_{1,79} = 3.45$, $p = 0.067$), but became significant when SOI was added to the analysis ($F_{1,75} = 4.03$, $p < .05$) despite SOI not interacting with altruism ($F < 1$). Altruism had no impact on preferences for dates with or without SOI in the analysis ($F_{1,78} = 2.17$ and $F_{1,75} = 2.12$, respectively, both n.s.), and SOI did not interact with altruism on preferences for dates ($F < 1.$). Female participants were more likely to lend money to altruistic males ($F_{1,79} = 5.53$, $p < 0.05$), or enjoy working with them ($F_{1,78} = 5.67$, $p < 0.05$), but had no preference for platonic friendships ($F_{1,75} = 1.61$, n.s.). Altruism had no effect on the physical or sexual attractiveness of the male targets (both Fs < 1, n.s.).

5.3.2(3) Effects of Altruism on Desirability of Females. There was no significant effect of altruism on men's willingness to have long-term relationships or dates with the target females, whether SOI was present in the analysis ($Fs_{1,79}$ = 1.23 and 1.69, respectively, both n.s., see Figure 5.2 & Table 5.3) or absent ($Fs_{1,70}$ = 0.48 and 1.82, respectively, both n.s.), and SOI did not interact with altruism to affect either variable (both Fs < 1). There was no effect of altruism on men's likelihood of lending money to the target females or enjoying working with them ($Fs_{1,79}$ = 1.65 and 0.26, both n.s.). However, altruism did increase men's reported willingness to have platonic friendships with the female targets ($F_{1,79} = 4.57$, $p <$ 0.05). Altruism had no effect on the physical or sexual attractiveness of female targets (both Fs < 1, n.s.).

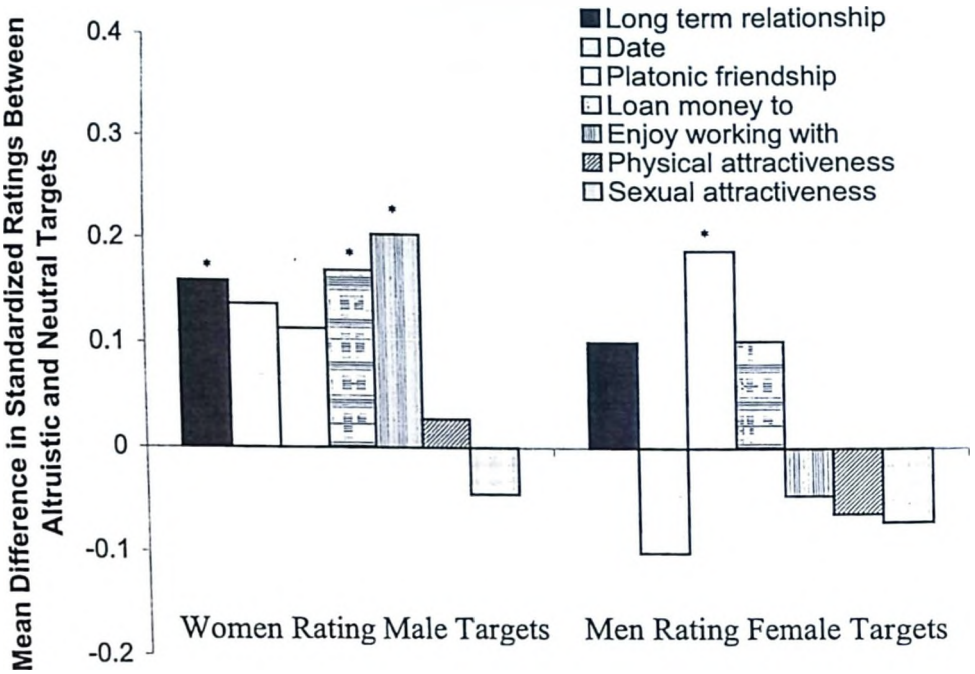


Figure 5.2: The effects of other-reported altruism (Experiment Two) on the desirability of male and female targets for long term relationships (black bars), dates (grey bars), platonic friendships (white bars), loans (horizontally striped bars), working partnerships (vertically striped bars), and the effects of altruism on physical attractiveness (diagonally striped bars) and sexual attractiveness (dotted bars). Ratings were standardized according to the mean and standard deviation of each picture on each variable. * $p < 0.05$

Table 5.3: Participants' standardized mean ratings (and standard errors) of target men and women on the target's desirability and attractiveness in Experiment Two. Ratings represent standard deviations from the sex-specific mean for each photograph on each variable.

	Target Seeking Short-term		Target Seeking Long-term	
	Neutral	Altruist	Neutral	Altruist
Long-Term Relationship				
Male Target	-0.34 (0.10)	-0.08 (0.11)	0.18 (0.12)	0.24 (0.11)
Female Target	-0.12 (0.11)	-0.06 (0.11)	0.03 (0.11)	0.16 (0.12)
Date				
Male Target	-0.21 (0.10)	-0.01 (0.12)	0.08 (0.11)	0.15 (0.12)
Female Target	0.05 (0.10)	-0.15 (0.11)	0.05 (0.11)	0.05 (0.12)
Platonic Friendship				
Male Target	-0.14 (0.11)	0.02 (0.11)	0.05 (0.12)	0.12 (0.11)
FemaleTarget	-0.06 (0.11)	0.14 (0.10)	-0.13 (0.12)	0.05 (0.13)
Work Partnership				
Male Target	-0.18 (0.08)	0.06 (0.08)	-0.04 (0.08)	0.14 (0.08)
FemaleTarget	-0.01 (0.12)	-0.03 (0.12)	0.05 (0.11)	-0.01 (0.10)
Loan Money				
Male Target	-0.24 (0.10)	-0.03 (0.11)	0.07 (0.11)	0.20 (0.12)
FemaleTarget	-0.11 (0.11)	-0.03 (0.11)	0.00 (0.11)	0.13 (0.12)
Physical Attractiveness				
Male Target	-0.08 (0.11)	0.00 (0.11)	0.07 (0.12)	0.04 (0.11)
Female Target	0.10 (0.10)	-0.08 (0.11)	0.03 (0.11)	-0.05 (0.13)
Sexual Attractiveness				
Male Target	-0.05 (0.11)	-0.01 (0.12)	0.11 (0.11)	-0.02 (0.11)
FemaleTarget	0.10 (0.10)	-0.01 (0.11)	-0.03 (0.11)	-0.06 (0.12)

5.3.2(4) Effects of Relationship Type Sought. Women rated the long-term-seeking targets more favourably than the short-term-seeking targets with respect to their stated willingness to have long-term relationship or dates with the targets or loan them money (all Fs > 7.5, all ps < 0.01), but not with respect to their willingness to have platonic friendships or work partnerships (both Fs < 2.5, both n.s.). Relationship Type did not affect men's preferences on any of these variables (all Fs < 3, all n.s.). Relationship type had no effect on either physical or sexual attractiveness for either sex (all Fs < 1.5).

5.3.2 Experiment Two Discussion

Experiment Two showed that women preferred the altruistic men to the neutral men for long-term relationships and were more likely to enjoy working with or loan money to the former. Other-reported generosity had no effect on men's preferences. This replicates the main findings of Experiment One by showing that women were attracted to altruism in some types of relationships, but men were not. This is important because different examples of altruism were used in the two studies, and it shows that this general pattern holds for both self-reported and other-reported generosity.

Experiment Two's results differed slightly from those of Experiment One. Females in Experiment One rated altruistic males as being more sexually attractive than neutral males and they preferred the former for dates and platonic friendships, but these effects were not found in Experiment Two. It is possible

that in Experiment One, women perceived the altruistic targets as having higher skills or abilities than the control targets, thus inspiring attraction even for shorter relationships such as dates. In Experiment Two, skill level was better controlled for, and this could have eliminated the difference in sexual attractiveness and women's preference for altruism in dates. Women may be attracted to altruists for long-term relationships (consistent with Dunbar & Kelly, 2001) regardless of the type of altruism because it signals character, but are only attracted to altruists for short relationships when the altruism also signals abilities. If the perceived skill level of altruists (relative to controls) was indeed the crucial difference between Experiments One and Two, then my findings would support that hypothesis, but further work is clearly needed on this topic. Men were more willing to have platonic friendships with the altruistic women than the neutral women, but otherwise their lack of preference for either type of female replicated the results of Experiment One.

Women in both experiments were more willing to have dates or long-term relationships with males who sought long-term relationships than with males who sought short-term relationships. This result is similar to the results of Experiment One, and is not surprising given that many people would presumably want their potential partners to be open to commitment. The subtlety of the manipulation of Relationship Type in Experiment Two allows us to generalize this finding farther than was possible with Experiment One because the targets would not have come across as being as sleazy as they might have in Experiment One. The increased

subtlety of the manipulation is likely responsible for the fact that men had no preference for long-term-seeking targets whereas women did, given that women are more sensitive to cues of commitment than men are (Buss & Schmitt, 1993). The more subtle manipulation could also explain the finding that a target's relationship type did not affect women's preferences for platonic friendships or working partnerships in Experiment Two, because the short-term-seeking targets did not come across as poorly as in Experiment One. Commitment is not as much of an issue for women in relationships where there is less risk of exploitation, such as non-romantic relationships, so it is not surprising that it had no effect on those types of relationships. It is interesting that the type of relationship sought by targets affected people's willingness to loan money to the targets, given that creditor/debtor relationships are clearly a type of relationship where exploitation (i.e. non-repayment) is a potential problem. Women in both experiments preferred to loan money to generous people, and this supports studies showing that (at least some) humans are good at detecting instances of altruism (Brown & Moore, 2000), and will tend to reward or trust generous people more than stingy people (Barclay, 2004, Chapter 2; Milinski et al., 2002; Wedekind & Milinski, 2000).

5.4 General Discussion

Together, these two studies show that women were attracted to altruistic men, in that descriptions of generous tendencies increased the desirability of the men in the descriptions. Furthermore, this effect was not simply "nice guys"

versus "jerks", because each altruistic description was paired with a neutral version rather than a negative version. Thus, this work expands and improves upon other studies on attractiveness (e.g. Jensen-Campbell et al., 1995; Kelly & Dunbar, 2001; Urbaniak & Kilmann, 2003) by using multiple descriptions with proper controls to demonstrate positive effects of altruism on attractiveness (in women's preferences, at least). Although these findings seem to contradict the popular wisdom that women do not want to date "nice guys", it is currently unknown whether real-life altruistic men tend to be less attractive than other men, such that they actually are less desirable overall. Unattractive men could use generosity or compassion to compensate for a lack of attractive qualities such as athleticism, courage, or physical attractiveness (which are possessed by stereotypical "bad boys") in order to make the best of a bad situation. Males with other attractive qualities might not signal via altruism if signaling those other qualities pays off better per unit of effort. Thus, even if altruistic acts can increase the desirability of any male, they might tend to be performed more often by less attractive men with fewer desirable traits, thereby creating the popular assumption that women do not prefer altruistic men. One might expect brave and athletic altruists to be the most desirable males, and Farthing (2005) indeed found that both sexes (but especially women) preferred heroic physical risk-takers to non-risk-takers. This might explain some contemporary women's apparent fascination with firemen (or at least firemen calendars), because firemen are expected to take physical risks in order to rescue others.

If altruism increases a male's desirability, then this can help account for the existence of altruistic displays by males in certain contexts. If males who display generosity tend to receive more (or better quality) partnerships than they would have if they had not made such displays, then this gives them an incentive for generous displays. If this also occurred in ancestral environments, then sexual selection could have selected for psychological mechanisms that increased the likelihood of generosity in ancestral men. It is important to note that this is complementary to, and not mutually exclusive with, a socialization account for the presence of altruism. Social rewards will increase or decrease the likelihood that any given man signals character via generosity, and female attention is a powerful reward for males.

The two studies suggest that men have no strong preference for or against generosity in women. This is consistent with the hypothesis that women are more concerned with good character in mates than men are. Men may have focused more on other aspects of the targets such as attractiveness and the relationship type sought by the targets, rather than generosity and compassion. Thus, women do not appear to gain mating benefits for being altruistic (in contrast to men), so it appears that mating benefits are not part of the reason for the existence of female altruism.

The present study used simulated dating advertisements to measure mate preferences. Several researchers have done content analyses of real lonely hearts advertisements to investigate mating strategies and preferences (e.g. Oda, 2001;

Thiessen, Young, & Burroughs, 1993; Wiederman, 1993). However, very few real advertisements explicitly mention altruistic tendencies or request them, making it infeasible to measure preferences for altruism by analyzing the content of existing ads. Strassberg and Holty (2003) created experimental personal advertisements and measured the hit rates of different types of ads. Future studies could use a similar procedure to test whether the current findings generalize to real-life mating contexts and further examine whether altruism increases a person's desirability.

Chapter 6: General Discussion

6.1 Costly Signaling via Altruism

6.1.1 Hypotheses and Support

Several researchers have argued that altruism may be a way of signaling cooperative intent (e.g. Gintis et al., 2001; Smith, 2003; Tessman, 1995). If this occurred sufficiently frequently in ancestral environments, then it could have provided a selective pressure shaping psychological mechanisms that regulate public generosity and enable people to respond to altruism with trust. I have tested some predictions derived from this idea, and the results support the notion that altruism can function as a signal.

If altruism functions as a costly signal of cooperative intent, then we might expect that people will be more generous in contexts where they can benefit from having a good reputation, and may even compete for such a reputation by increasing cooperation levels even more. I supported these predictions in Chapter 2, because participants did contribute more to public goods when they were informed about the subsequent trust games. Furthermore, monetary contribution levels were best maintained when there was an opportunity to compete for potential reputational benefits. Participants seemed to respond to generosity as if it were a signal of cooperative intent, because people who contributed more to public goods were trusted more than people who contributed less. In Chapter 4, I replicated the latter result, and also showed that people entrusted more money to those who had paid to punish free-riders than to those who had not (or who had

punished less) after they had repeated experience with free-riders. This supports the hypothesis that people respond to justified punishment as if it were a signal of cooperative intent, and this hypothesis is further supported by the finding that it was only justified punishers who were trusted. Unjustified punishment does not signal a dislike or intolerance of unfairness like justified punishment does, so it is unsurprising that unjustified punishers were trusted less. I did not test whether people are more punitive when they can gain a reputation for punishment (i.e. whether they *use* punishment as a signal). However, Kurzban et al. (2004) did test this, and found that people were less willing to perform third-party punishment when their decisions were completely anonymous than when they were not.

I also argued that costly signaling theory predicts that punishment can signal an unwillingness to tolerate unfairness or transgressions against oneself. From this I predicted that people who are more concerned with demonstrating toughness will be more likely to punish free-riders. Men are probably more concerned with signaling such traits than women are because men have a greater tendency than women to deter transgressions with a credible use of force (Cohen et al., 1996; Daly & Wilson, 1988). Indeed, I found that men tended to punish free-riders more than women did in Chapters 3 and 4. Punishing free-riders may be a socially acceptable way to signal toughness.

I predicted that those who have a greater ability to be generous or to punish will be more likely to actually do so, and high social status is one thing that increases a person's ability to be altruistic or punitive. In Chapter 3, I tested

the prediction that people who were accorded high status within a group would be more likely to contribute to public goods and to punish free-riders. This prediction was not confirmed, but this failure may have been due to a failure to adequately alter perceptions of status rather than any weakness in the hypothesis. The data did not support my predictions, but did provide some evidence that participants may be affected by reputational concerns in that physical proximity to the experimenter had a significant effect on participants' levels of cooperation and punishment. Although such proximity did not affect the anonymity of participants, it could have triggered reputational concerns by affecting their perceptions of anonymity. Haley and Fessler (2005) and Burnham and Hare (*in press*) have found that subtle cues of non-anonymity (namely human-like eyes or stylized eyespots on a computer screen) significantly increased people's altruism even though participants supposedly knew that their decisions were completely anonymous. Physical proximity to the experimenter could have caused a similar effect in my experiment.

If altruism signals character or abilities, then we would expect that altruists are more desirable as mating partners than non-altruists. In Chapter 5, women reported being more willing to have long-term relationships with men who were depicted as altruists than with men who were not. This supports the hypothesis that women should be sensitive to signals of character, and suggests that men might benefit from using altruism to send such signals.

6.1.2 Responder Behaviour: Trust or Reward?

Regardless of whether cooperation functioned as a signal in my experiments (strategically or unconsciously), it certainly seems that people responded to cooperation as if it were a signal. In Chapters 2 and 4, people who contributed more to public goods were trusted more than those who contributed less, and in Chapter 4 those who provided altruistic (justified) punishment were trusted more than those who did not (or did so to a lesser extent). In those chapters, participants in the Truster role could benefit from entrusting money to Responders, because Responders might return more money than was sent, but there was a chance that Responders would not return any money, such that the Trusters would be worse off for having sent any money.

Milinski et al. (2002a, b) showed that people tend to reward those who contribute to public goods even when no money could be returned, which suggests that my "trust" game results might be explained by rewarding of high contributors rather than by trust per se. However, I have several reasons for thinking that these results are more likely to represent trust of high contributors (because altruism signal cooperative intent) rather than merely rewarding high contributors (as predicted by reciprocity theory). First, participants in Milinski et al.'s experiments (2002a, b) did in fact have a strategic reason to reward cooperators: to motivate future contributions to public goods. In my experiments, there was no such strategic reason to reward cooperators, so the desire to reward

would be reduced. Secondly, the amounts entrusted were much greater than the amounts that people spent to reward cooperators in other experiments (Clark, 2002; Sefton et al., 2002).[11] Thirdly, people entrusted non-zero amounts to free-riders despite the fact that they often punished free-riders. It does not make sense that people would punish low contributors only to reward them immediately afterwards, whereas it is conceivable that they might still have some (albeit little) faith that the low contributors would return some money. Fourthly, the people who contributed more to the public goods did not entrust more than others who contributed less, which speaks against an indirect reciprocity account because one would expect the cooperators to spend more on rewarding. Fifthly, in Chapter 4, participants entrusted more money to punishers but did not return more money to them. The fact that Responders in the Trust Game did not return more money to punishers suggests that players trusted punishers more than non-punishers but were not motivated to reward them more than non-punishers. Sixthly, with direct or indirect reciprocity, one only needs to be as altruistic as others in one's group in order to be considered cooperative, and one does not necessarily need to be more altruistic than others. With costly signaling, competition might not always occur, but is likely to occur when each signaler needs to send a stronger signal than competitors in order to be chosen over others as a mate or partner. Thus, the

[11] I cannot directly compare amounts with Milinski et al.'s (2002a, b) studies, because their participants made binary "yes" or "no" decisions about rewarding, so we do not know how much they would reward if their decisions had been unconstrained.

evidence for competitive altruism in Chapter 2 is more consistent with a costly signaling model of altruism than just indirect reciprocity.

6.1.3 Multiple Effects of Altruism

Although I believe that these results support a costly signaling account for altruism better than they support an indirect reciprocity account, the hypothesized benefits of altruism are not necessarily mutually exclusive. For example, people "showing off" with altruism could receive more help when in need (indirect reciprocity), associate with other good cooperators (assortative interactions), have higher quality or more numerous mates (costly signaling of individual quality), be deferred to more often (costly signaling of individual quality), while at the same time be trusted more than others because they are deemed more trustworthy (costly signaling of cooperative intent). For other group members, giving aid to these altruists serves the function of keeping oneself in good standing with others (indirect reciprocity) and close enough to them to receive the benefits of their magnanimity (assortative interactions), mating with them serves the function of acquiring high quality mates and deferring to them allows one to avoid costly competition with a superior competitor (costly signaling of individual quality), and trusting them serves the function of interacting preferentially with those who are less likely to cheat a cooperative partner (costly signaling of cooperative intent).

That being said, anything that increases the benefits to the altruist can also reduce the effectiveness of generosity as a costly signal because the benefits of

the signal might outweigh the costs even for lower quality (or less cooperative) individuals. Thus, as the benefits of signaling via altruism increase, we would expect altruistic signals to become more costly in order to maintain their honesty. Berman (2003) and Sosis and Alcorta (2003) use similar arguments to explain the existence of costly signals of membership in certain religious groups; as the benefits of group membership increase, religious groups will impose greater costs and restrictions upon members in order to deter potential free-riders. Thus, a costly signaling account of altruism clearly predicts an escalation in the degree of generosity displayed when the potential benefits of altruistic signaling increase. Moreover, costly signaling and assortative interactions both predict an escalation of generosity in response to increased generosity by others, in that one should be motivated to signal one's higher quality than that of competitors or be more cooperative in order to pair up with the most generous partners. Thus, the different theories do make different predictions about altruism. While the processes that select for altruistic sentiment may all be operating at the same time, some may be more important in some situations than others, and this may have been true in the ancestral environments that shaped the psychology of cooperation and punishment.

If altruism does function as a costly signal of cooperative intent, then we might expect that receivers should be sensitive to the costs and benefit that another person experiences when sending a signal of cooperation, in order to determine whether the signal is honest or not. When many people perceive the

signal, the potential benefits for signaling cooperative intent are greater than when only one person perceives the signal because many people may start to trust the signaler. Thus, we might predict that perceivers will be sensitive to the number of people who see the signal. When audiences are bigger, signals of commitment must be larger in order to be considered honest because the potential benefits of such signals are greater. Future studies could investigate the effects of audience size on the trustworthiness of altruism.

6.2 Debates Over Group Selection

6.2.1 Implications for Group Selection

Proponents of group selection (e.g. Boyd & Richerson, 2002; Sober & Wilson, 1998; Henrich & Fehr, 2003) have sometimes argued that individual-level benefits cannot account for the existence of altruistic behaviour (such as the provision of public goods) because altruists often cannot target their generosity toward specific individuals. Thus, they claim that group selection was likely a significant force in the evolution of human cooperation. This argument assumes that people treat cooperation in collective action projects and cooperation in dyadic relationships as completely separate phenomena, as if to say, "What happens in collective action relates only to collective action, what happens in dyadic relationships relates only to dyadic relationships, and never the twain shall meet." However, life is not a series of separate one-shot situations, and a person's reputation in one type of situation (like a collective action problem) can carry

over into other situations (such as dyadic interactions). People who provide public goods may be disadvantaged relative to free-riders at the moment when they provide the public good, but receive subsequent reputational benefits that compensate for this disadvantage. This thesis presents evidence that people do receive reputational benefits from providing public goods, and these reputational benefits can translate into tangible benefits.

Chapter 2 showed that those who contribute to public goods are trusted more in subsequent dyadic interactions than those who contribute less. Chapter 4 replicated this finding, and demonstrated that punishers also receive social benefits when they impose sanctions on uncooperative individuals. These chapters showed that people entrusted more money to public goods providers and justified punishers, which shows that altruists can receive tangible benefits for their behaviour. Finally, Chapter 5 showed that altruistic men were more appealing as long-term romantic partners, and suggested that altruistic men might thus attract more (or higher quality) women for romantic relationships. Thus, it seems clear that people can receive individual-level benefits from altruistic behaviour. If altruism and justified punishment are signals of cooperative intent, then these benefits will not be subject to the second-order free-riding problem because others will cooperate with altruists and justified punishers because it is in their best interest to do so. If similar benefits were accrued in ancestral environments, then this could have selected for the psychological mechanisms that modulate altruistic and punitive behaviour, regardless of any group-level advantages. The presence of

individual-level benefits implies that cooperative and punitive sentiments do not function solely to benefit one's group, and the effect on the group could simply be a byproduct of mechanisms that have evolved to increase one's own fitness within a group.

This is not to say that group selection cannot account for the spread of cooperative sentiments between groups. Individual-level adaptations are more likely to spread between groups if they also happen to be beneficial for one's group than if they are detrimental to one's group (Boyd & Richerson, 2002). However, group-beneficial acts that are also individually beneficial are more likely to be group selected than group-beneficial acts that are individually costly, because group and individual-level selection will not be in conflict when the acts happen to be individually beneficial. Individual-level selection helps determine the prevalence of cooperative sentiments within groups, and then group selection could act on the group-level differences that result. Groups in which altruism is individually beneficial will have more altruism than those in which it is individually costly, because there will be within-group selection for altruism in the former and within-group selection against altruism in the latter. Because of the resulting differences in cooperation, the former groups would also have higher fitness than the latter groups. Thus, group selection is more likely to select for cooperative sentiments that bring individual-level *and* group-level benefits than cooperative sentiments that bring only group-level benefits, simply because the

former will be more prevalent within groups and thus cause greater between-group differences.

6.2.2 Laboratory Environments and Proximate Mechanisms

The present results suggest that altruism and altruistic punishment can be used and interpreted as signals of cooperative intent. However, in most studies using public goods games, Trust Games, and Prisoner's Dilemmas, participants make anonymous decisions in order to reduce the possibility that they will be influenced by reputational concerns. In anonymous, one-shot interactions, participants cannot benefit from acquiring a good reputation, yet many participants continue to cooperate with non-kin at non-zero levels despite apparently comprehending the lack of reputational opportunities. Because of such findings, some researchers (Fehr et al., 2002; Fehr & Henrich, 2003; Gintis, Bowles, Boyd, & Fehr, 2003) claim that inclusive fitness, reciprocal altruism and costly signaling cannot account for the levels of altruism found in humans, and so cooperative sentiment must have been selected for by between-group selection in ancestral environments.

However, such reasoning is partly based on confusion between the proximate and ultimate causes of cooperative behaviour (despite claims to the contrary, see Fehr et al., 2002; Gintis et al., 2003; Henrich & Fehr, 2003). If there have been past selective pressures favouring altruism, then natural selection would be expected to have selected for some sort of cognitive mechanisms or cooperative sentiments or decision rules that would regulate such behaviour (or

allow it to be learned). Once such mechanisms exist, they would function even if a person is in an unfamiliar situation such as a laboratory experiment, and *especially* if participants determine what is "appropriate" in such situations by comparing them to familiar situations outside the laboratory (Henrich et al., 2004). If people receive emotional rewards or positive feelings from cooperating or punishing non-cooperation (de Quervain et al., 2004), then those people are likely to receive those rewards or have those feelings whether they are in a laboratory or not. In fact, there have to be some sort of proximate mechanisms that regulate cooperation (even in the laboratory), unless we postulate that external forces cause this behaviour without affecting the brain in any recognizable way. Thus, a person may very well be altruistic or punitive because he/she enjoys being altruistic or punitive, even if he/she consciously believes that his/her actions are unknown to others. However, the presence of such a proximate mechanism says nothing about the selection pressures that would have caused such a mechanism to exist in the first place.

If one suggests that the mechanisms that cause cooperation in the laboratory are the same as those that cause it outside the laboratory, then one is not saying that the decision rules must be insensitive to information about anonymity or reputation. We might expect cooperative sentiments to be sensitive to real-world cues of the likelihood of being observed, such that they promote more cooperation in the presence of others. Consistent with this, Gächter and Fehr (1999), Rege and Telle (2004), and Hoffman, McCabe, Schachat, and Smith

(1994) found that people are less cooperative when their decisions are anonymous.

However, there is good reason to suspect that such sentiments will promote cooperation even if a person believes that he/she is not observed. Frank (1988) has argued that one's perceptions of anonymity can occasionally be wrong. If so, then a person might accidentally defect while not completely anonymous if he/she possessed a psychology that permitted defection under conditions of suspected anonymity. Such defections would likely hurt his/her reputation and undermine the cooperative image that he/she built. Frank provided a model showing that selection favours mechanisms that promote some degree of cooperation even under conditions of perceived anonymity, and the error rates do not have to be particularly high for this to be true. Thus, we should not expect perceptions of anonymity to completely eliminate all cooperative behaviour. If a person does receive cues of being observed (including but not limited to the conscious knowledge that one can acquire a reputation), then this would trigger reputation-enhancing mechanisms that raise one's cooperation above this baseline level. Haley and Fessler (2005) and Burnham and Hare (*in press*) have found that the presence of stylized eyespots or human-like eyes on a computer are sufficient to significantly increase people's level of altruism, even under conditions of complete anonymity. This suggests that the psychological adaptations that function to maintain reputation can be triggered even when people consciously know that they cannot acquire a reputation. Similarly, DeBruine (2002) and

Krupp, DeBruine and Barclay (2005) show that subtle cues of kinship affect trust and cooperation even in experimental games in which participants believe that they are interacting with unrelated strangers.

When people cooperate in anonymous one-shot interactions, they act in ways that are individually costly yet beneficial to others. Such behaviour may be adaptive outside the laboratory, but is apparently maladaptive inside the laboratory where interactions are anonymous. Proponents of group selection use this maladaptiveness to suggest that reputation-based models cannot account for such behaviour, so we should accept that group selection has been a major force in the evolution of altruistic sentiments (e.g. Fehr & Henrich, 2003). However, these arguments neglect the fact that such behaviour is maladaptive even from a group-selectionist perspective. Under group selection, altruism could be selected for whenever groups with relatively more altruists tend to out-reproduce or out-compete groups with relatively fewer altruists. However, there is no selection-relevant group within laboratory environments. Groups in laboratory environments are not competing against other laboratory groups, and group selection obviously cannot act directly on short-lived laboratory groups because group members rejoin their own social groups as soon as the experiment is over. Thus, group selection cannot account for altruism in laboratory experiments unless the selection-relevant group is the entire university community, and this is much bigger than the group sizes in which group selection is viable (Boyd et al., 2003). Furthermore, people will cooperate in anonymous games with people (or

what they believe are real people) in other parts of the country or even other countries (DeBruine, 2002; Eckel & Wilson, 2004; Yamagishi et al., *in press*), so if group selection were a significant factor in the evolution of cooperative sentiments then the "groups" upon which the between-group selection was acting would have to be the entire country or the whole human species. This is clearly implausible. When people cooperate in anonymous laboratory experiments, they are not only being cooperative when they cannot acquire a reputation from doing so, but they are also cooperating in a manner that could not have been selected for via group selection. Thus, I conclude that people are acting in ways that are clearly maladaptive when they cooperate in anonymous one-shot interactions, regardless of what selective pressures caused altruism to evolve and however such behaviour might have been adaptive in ancestral environments.

When altruistic behaviour is maladaptive in the laboratory under either group-selection or individual-level selection, it is illogical to use the fact of maladaptation to support theories of group selection over reputation-based theories. Group-selectionist and reputation-based theories both rely on people having preferences that are adaptive outside the laboratory but are not always adaptive in experiments or other rare "completely anonymous" interactions. By using findings from anonymous one-shot games to attack reputation-based theories, proponents of group selection are neglecting the fact that such findings do not support group selection either. Hence, there appears to be a severe flaw in a major argument used in support of group-selected altruistic behaviour in

humans. Based on this fact and based on the evidence I have discussed in support of reputation-based theories, I would suggest that the balance of experimental evidence now supports the notion that altruistic sentiments are the product of reputation-based individual-level selection (including indirect reciprocity and costly signaling) rather than group selection.

6.3 Potential Future Directions

I have presented evidence supporting the notion that altruism can function as a signal of cooperative intent. Costly signaling theory has only recently been applied to human behaviour, and particularly to human cooperation. This is an area that is ripe for theoretical and empirical work. We are still in need of mathematical models to formalize the conditions under which we might expect altruism to be used as a costly signal of cooperative intent, and what factors should affect how receivers respond to such signals. Future work can focus on the conditions under which people will compete to be more altruistic than others, and the effects of audience size and characteristics. Furthermore, altruists probably receive benefits other than trust for their actions, so future work can investigate what other benefits are accrued by generous people. For example, punishers are likely to be feared or deferred to but not necessarily liked, so one might be interested in testing whether this is indeed the case. Finally, it would be useful to examine the particular emotions and decision rules involved in cooperation and justified punishment, and how those mechanisms develop within a person's

lifetime. Such work could contribute even further to our understanding of human cooperation.

References

Albert, M., Güth, W., Kirchler, E., & Maciejovsky, B. (2002). Are we nice(r) to nice(r) people? An experimental analysis. *Discussion Paper 2002-15*, Max Planck Institute for Research into Economics Systems, Strategic Interaction Group, Jena, Germany. ftp://papers.mpiew-jena.mpg.de/esi/discussionpapers/2002-15.pdf

Alcock, J. (1993). *Animal Behavior: An Evolutionary Approach*. Sunderland, MA: Sinauer Associates.

Allan, S., & Gilbert, P. (1995). A social comparison scale: psychometric properties and relationship to psychopathology. *Personality and Individual Differences*, *19(3)*, 293-299.

Alexander, R. D. (1987). *The Biology of Moral Systems*. New York, NY: Aldine de Gruyter.

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

Atran, S. (2001). A cheater-detection module? Dubious interpretations of the Wason Selection Task and logic. *Evolution and Cognition*, *7*, 1-7.

Axelrod, R. (1984). *The Evolution of Cooperation*. New York, NY: Basic Books.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 1390-1396.

Ball, S., & Eckel, C. (1996). Buying status: experimental evidence on status in negotiation. *Psychology and Marketing*, *13*, 381-405.

Ball, S., & Eckel, C. (1998). The economic value of status. *Journal of Socio-Economics*, *27*, 495-514.

Ball, S., Eckel, C., Grossman, P. J., & Zame, W. (2001). Status in markets. *Quarterly Journal of Economics*, *116*, 161-188.

Barclay, P. (*in preparation*). Self-perceived social status and behaviour in experimental social dilemmas.

Barclay, P., & Lalumière, M. L. (*in press*). Do people differentially remember cheaters? *Human Nature*.

Barr, A. (2001) Social dilemmas and shame-based sanctions: Experimental results from rural Zimbabwe. *Working Paper WPS/2001.11*, Centre for the Study of African Economics, University of Oxford, U.K.

Barr, A., & Kinsey, B. (2002). Do men really have no shame? *The Centre for the Study of African Economies Working Paper Series*. Working Paper 164. http://www.bepress.com/csae/paper164/

Barrett, L., Henzi, S. P., Weingrill, T., Lycett, J. E., & Hill, R. A. (2000). Female baboons do not raise the stakes but they give as good as they get. *Animal Behaviour, 59*, 763-770.

Berg, J., Dickhaut, J., & McCabe, K. 1995 Trust, Reciprocity and Social History. *Games & Economic Behavior, 10*, 122-142.

Berman, E. (2003). Hamas, Taliban and the Jewish Underground: an economist's view of radical religious militias. *National Bureau of Economic Research (NBER) Working Paper Series*. Working Paper 10004. http://www.nber.org/papers/w10004

Betzig, L. (1988). Redistribution: equity or exploitation? In *Human Reproductive Behaviour: A Darwinian Perspective* (Eds. L Betzig, M. Borgerhoff Mulder, & P. Turke), pp. 49-63. Cambridge, UK: Cambridge University Press.

Betzig, L. L., & Turke, P. W. (1986). Food sharing on Ifaluk. *Current Anthropology, 27*, 397-400.

Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology, 3*, 27-52.

Bliege Bird, R., Bird, D. W., Smith, E. A., & Kushnik, G. C. (2002). Risk and reciprocity in Meriam food sharing. *Evolution and Human Behavior, 23*, 297-321.

Bliege Bird, R., & Smith, E. A. (2005). Signaling theory, strategic interaction, and symbolic capital. *Current Anthropology, 46*, 221-248.

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.

Bolton, G. E., Katok, E., & Ockenfels, A. (*in press*). Cooperation among strangers with limited information about reputation. *Journal of Public Economics*.

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

Boone, J. L. (1998). The evolution of magnanimity: when is it better to give than to receive? *Evolution and Human Behavior, 9*, 1-21.

Boone, J. L., & Kessler, K. L. (1999). More status or more children? Social status, fertility reduction, and long-term fitness. *Evolution and Human Behavior, 20*, 257-277.

Booth, A., Shelley, G., Mazur, A., Tharp, G., Kittok, R. (1989). Testosterone, and winning and losing in human competition. *Hormones and Behavior, 23*, 556-571.

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *PNAS, 100*, 3531-3535.

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology, 13*, 171-195.

Boyd, R., & Richerson, P. J. (2002). Group beneficial norms can spread rapidly in a structured population. *Journal of Theoretical Biology, 215*, 287-296.

Brandt, H., Hauert, C., & Sigmund, K. (2003). Punishment and reputation in spatial public goods games. *Proceedings: Biological Sciences, 270*, 1099-1104.

Brown, D. E. (1991). *Human Universals*. New York, NY: McGraw-Hill.

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.

Burger, J.M., & Cosby, M. (1999). Do women prefer dominant men? The case of the missing control condition. *Journal of Research in Personality, 33*, 358-368.

Burnham, T., & Hare, B. (*in press*). Engineering cooperation: does involuntary neural activation increase public goods contributions? *Human Nature*.

Buss, D.M., Abbot, M., Angleitner, A., Asherian, A, Biaggio, A., and 45 other co-authors (1990). International Preferences in Selecting Mates. *Journal of Cross-Cultural Psychology, 21*, 5-47.

Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review, 100*, 204-232.

Caldwell, M. D. (1976). Communication and sex effects in a five-person Prisoner's Dilemma game. *Journal of Personality and Social Psychology, 33*, 273-280.

Chagnon, N. (1988). Life histories, blood revenge, and warfare in a tribal population. *Science, 239*, 985-992.

Chan, K. S., Godby, R., Mestelman, S., & Muller, R. A. (1996). Spite, guilt and the voluntary provision of public goods when income is not distributed equally. *Canadian Journal of Economics, 29 (Special Issue)*, S605-609.

Chan, K. S., Godby, R., Mestelman, S., & Muller, R. A. (1997). Equity theory and the voluntary provision of public goods. *Journal of Economic Behavior and Organization, 32*, 349-364.

Chan, K. S., Mestelman, S., Moir, R., Muller, R. A. (1996). The voluntary provision of public goods under varying income distributions. *Canadian Journal of Economics, 96*, 54-69.

Chang, A., & Wilson, M. (2004). Recalling emotional experiences affects performance on reasoning problems. *Evolution and Human Behavior, 25*, 267-276.

Cheng, P. W., & Holyoak, K. J. (1989). On the natural selection of reasoning theories. *Cognition, 33*, 285-313.

Clark, J. (2002). Recognizing large donations to public goods: An experimental test. *Managerial and Decision Economics, 23*, 33-44.

Clutton-Brock, T. H., & Parker, G. A. (1995). Punishment in animal societies. *Nature, 373*, 209-216.

Cohen, D., Nisbett, R. E., Bowdle, B. F., & Schwarz, N. (1996). Insult, aggression, and the southern culture of honor: an "experimental ethnography". *Journal of Personality and Social Psychology, 70*, 945-960.

Commins, B., & Lockwood, J. (1979). The effects of status differences, favoured treatment and equity on intergroup comparisons. *European Journal of Social Psychology*, *9*, 281-289.

Connor, R. C. (1996). Partner preferences in by-product mutualisms and the case of predator inspection in fish. *Animal Behaviour*, *51*, 451-454.

Cordell, J., & McKean, M.A. (1992). Sea Tenure in Bahia, Brazil. In *Making the Commons Work: Theory, Practice, and Policy* (ed. D. W. Bromley), pp. 183-205. San Francisco: ICS Press.

Cosmides, L., & Tooby, J. (1992). Cognitive Adaptations for Social Exchange. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (eds. J. Barkow, L. Cosmides, & J. Tooby), pp.163-228. New York: Oxford University Press.

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, *46*, 260-281.

Daly, M., & Wilson, M. (1983). *Sex, Evolution, and Behavior*. Belmont, CA: Wadsworth Publishing Company.

Daly, M., & Wilson, M. (1988). *Homicide*. New York, NY: Aldine de Gruyter.

Davis, D.D., Holt, C.A. (1993). *Experimental Economics*. Princeton: Princeton University Press.

Dawes, R. M., McTavish, J., & Shaklee, H. (1977). Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of Personality and Social Psychology*, *35*, 1- 11.

Dawes, R. M., & Messick, D. M. (2000). Social Dilemmas. *International Journal of Psychology*, *35*, 111-116.

Dawkins, R. (1976). *The Selfish Gene*. Oxford, UK: Oxford University Press.

DeBruine, L. (2002). Facial resemblance enhances trust. *Proceedings: Biological Sciences*, *269*, 1307-1312.

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

Drucker, P., & Heizer, R. F. (1967). *To Make My Name Good: A Reexamination of the Southern Kwakiutl Potlatch*. Berkeley, CA: University of California Press.

Dugaktin, L. A. (2004). *Principles of Animal Behavior*. New York, NY: W. W. Norton & Company.

Dwyer, P. D., & Minnegal, M. (1993). Are Kubo hunters 'show offs'? *Ethology and Sociobiology, 14*, 53-70.

Dwyer, P. D., & Minnegal, M. (1997). Sago games: cooperation and change among sago producers of Papua New Guinea. *Evolution and Human Behavior, 18*, 89-108.

Eckel, C. C., & Wilson, R. K. (2004). Is trust a risky decision? *Journal of Economic Behavior and Organization, 55*, 447-465.

Ellis, L. (1993). Operationally defining social stratification in human and nonhuman animals. In *Socioeconomic Inequality, Volume 1: A Comparative Biosocial Analysis* (Ed. L. Ellis), pp. 15-36. Westport, CT: Praeger.

Ensminger, J. (2004). Market integration and fairness: Evidence from Ultimatum, Dictator, and Public Goods Experiments in East Africa. In *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies* (Eds. J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, & H. Gintis), pp 356-381. Oxford, UK: Oxford University Press.

Farthing, G. W. (2005). Attitudes toward heroic and nonheroic physical risk takers as mates and friends. *Evolution and Human Behavior, 26*, 171-185.

Faurie, C., Pontier, D., Raymond, M. (2004). Student athletes claim to have more sexual partners than other students. *Evolution and Human Behavior, 25(1)*, 1-8.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*, 785-791.

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences, 8*, 185-190.

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation and the enforcement of social norms. *Human Nature, 13*, 1-25.

Fehr, E. & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*, 980-994.

Fehr, E. & Gächter, S. (2002) Altruistic punishment in humans. *Nature, 415*, 137-140.

Fehr, E., & Henrich, J. (2003). Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism. In *Genetic and Cultural Evolution of Cooperation* (Ed. P. Hammerstein), pp. 55-82. Cambridge, MA: MIT Press.

Fershtman, C., & Weiss, Y. (1998). Social rewards, externalities and stable preferences. *Journal of Public Economics, 70*, 53-73.

Fessler, D. M. T. (2002). Windfall and socially distributed willpower: the psychocultural dynamics of rotating savings and credit associations in a Bengkulu village. *Ethos, 30*, 25-48.

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

Fodor, J. (2000). Why we are so good at catching cheaters. *Cognition, 75*, 29-32.

Frank, R. H. (1988). *Passions Within Reason.* New York, NY: Norton.

Gächter, S., & Fehr, E. (1999). Collective action as social exchange. *Journal of Economic Behavior and Organization, 39*, 341-369.

Gächter, S., & Herrmann, B. (2004). Norms of cooperation among urban and rural dwellers: experimental evidence from Russia. Talk presented at the 16th Annual Meeting of the Human Behavior and Evolution Society, Free University of Berlin, Germany (July 2004).

Gangestad, S.W., & Simpson, J.A. (2000). The evolution of human mating: trade-offs and strategic pluralism. *Behavioral and Brain Sciences, 23*, 573-644.

Gaulin, S. J. C., & McBurney, D.H. *Psychology: An Evolutionary Approach.* Upper Saddle River, NJ: Prentice Hall.

Ginsburg, H. J., & Miller, S. M. (1981). Altruism in children: A naturalistic study of reciprocation and an examination of the relationship between social dominance and aid-giving behavior. *Ethology and Sociobiology, 2*, 75-83.

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology, 206*, 160-179.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior, 24*, 153-172.

Gintis, H., Smith, E. A., & Bowles, S. (2001). Cooperation and costly signaling. *Journal of Theoretical Biology, 213*, 103-119.

Goldman, I. (1937). The Kwakiutl Indians of Vancouver Island. In *Cooperation and Competition Among Primitive People* (Ed. M. Mead) pp. 180-209. Boston, MA: Beacon Press.

Grayson, D. K. (1993). Differential mortality and the Donner Party disaster. *Evolutionary Anthropology, 2*, 151-159.

Gunnthorsdottir, A., Houser, D., McCabe, K., & Ameden, H. (2000). Excluding free-riders improves reciprocity and promotes the private provision of public goods. Working Paper.

Gurven, M., Allen-Arave, W., Hill, K., & Hurtado, A. M. (2000). "It's a Wonderful Life": signaling generosity among the Ache of Paraguay. *Evolution and Human Behaviour, 21*, 263-282.

Gurven, M., Allen-Arave, W., Hill, K., & Hurtado, A. M. (2001). Reservation food sharing among the Ache of Paraguay. *Human Nature, 12*, 273-297.

Gurven, M., Hill, K., Kaplan, H., Hurtado, A., & Lyles, R. (2000). Food transfers among Hiwi foragers of Venezuala: tests of reciprocity. *Human Ecology, 28*, 171-218.

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues enhance generosity in an anonymous economic game. *Evolution and Human Behavior, 26*, 245-256.

Hames, R. (1987). Garden labor exchange among the Ye'kwana. *Ethology and Sociobiology, 8*, 259-284.

Hamilton, W. D. (1964). The genetical evolution of social behaviour (I & II). *Journal of Theoretical Biology, 7*, 1-52.

Hammerstein, P. (2003). Why is reciprocity so rare in animals? A Protestant appeal. In *Genetic and Cultural Evolution of Cooperation* (Ed. P. Hammerstein), pp. 83-94. Cambridge, MA: MIT Press.

Harbaugh, W. T. (1998). What do donations buy? A model of philanthropy based on prestige and warm glow. *Journal of Public Economics, 67*, 269-284.

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

Harpending, H. (1998). Comment on D. S. Wilson's "Hunting, sharing, and multilevel selection: the tolerated-theft model revisited". *Current Anthropology, 39*, 88-89.

Hauert, C., Haiden, N., & Sigmund, K. (2004). The dynamics of public goods. *Discrete and Continuous Dynamical Systems – Series B, 4*, 575-587.

Hauser, M. D., Chen, M. K., Chen, F., & Chuang, E. (2003). Give unto others: genetically unrelated cotton-top tamarin monkeys preferentially give food to those who altruistically give food back. *Proceedings: Biological Sciences, 270*, 2363-2370.

Hawkes, K. (1990). Why do men hunt? Benefits for risky choices. In *Risk and Uncertainty in Tribal and Peasant Economies* (ed. E. Cashdan), pp. 145-166. Boulder, CO: Westview Press.

Hawkes, K. (1991). Showing off: tests of an hypothesis about men's foraging goals. *Ethology an Sociobiology, 12*, 29-54.

Hawkes, K. (1993). Why hunter-gatherers work: an ancient version of the problem of public goods. *Current Anthropology, 34*, 341-361.

Hawkes, K., & Bliege Bird, R. (2002). Showing off, handicap signaling, and the evolution of men's work. *Evolutionary Anthropology, 11*, 58-67.

Hawkes, K., O'Connell, J. F., & Blurton Jones, N. G. (2001a). Hadza meat sharing. *Evolution and Human Behavior, 22*, 113-142.

Hawkes, K., O'Connell, J. F., & Blurton Jones, N. G. (2001b). Hunting and nuclear families: some lessons from the Hadza about men's work. *Current Anthropology, 42*, 681-709.

Hawley, P. H. (1999). The ontogenesis of social dominance: A strategy-based evolutionary perspective. *Developmental Review, 19*, 97-132.

Henrich, J., & Boyd, R. (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology, 208*, 79-89.

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, UK: Oxford University Press.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In search of Homo economicus: Behavioral experiments from 15 small-scale societies. *American Economic Review, 91*, 73-78.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R., Alvard, M., Barr, A., Ensminger, J., Smith Henrich, N., Hill, K., Gil-White, F. J., Gurven, M., Marlowe, F. W., Patton, J. Q., & Tracer, D. (*in press*). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*.

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior, 22*, 165-196.

Hill, K., & Kaplan, H. (1988). Tradeoffs in male and female reproductive strategies among the Ache: part 1. In *Human Reproductive Behaviour: A Darwinian Perspective* (Eds. L Betzig, M. Borgerhoff Mulder, & P. Turke), pp. 277-289. Cambridge, UK: Cambridge University Press.

Hoffman, E., McCabe, K., Schachat, K., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior, 7*, 346-380.

Jensen-Campbell, L.A., Graziano, W.G., & West, S.G. (1995). Dominance, prosocial orientation, and female preferences: Do nice guys really finish last? *Journal of Personality and Social Psychology, 68*, 427-440.

Kaplan, H., & Hill, K. (1985). Food sharing among Ache foragers: tests of explanatory hypotheses. *Current Anthropology, 26*, 223-246.

Karp, D., Jin, N., Yamagishi, T., & Shinotsuka, H. (1993). Raising the minimum in the minimal group paradigm. *The Japanese Journal of Experimental Social Psychology, 32*, 231-240.

Kelly, S., & Dunbar, R.I.M. (2001). Who dares, wins: Heroism versus altruism in women's mate choice. *Human Nature, 12*, 89-105.

Keser, C. (2003). Experimental games for the design of reputation management systems. *IBM Systems Journal, 42*, 498-506.

Ketelaar, T., Au, W. T. (2003). The effects of feelings of guilt on the behaviour of uncooperative individuals in repeated social bargaining games: An affect-as-information interpretation of the role of emotion in social interaction. *Cognition and Emotion, 17*, 429-453.

Kiyonari, T., & Barclay, P. (2005). Selective incentives for cooperation: second-order punishment vs. second-order reward. Talk presented at the 17th Annual Meeting of the Human Behavior & Evolution Society, Austin, Texas (June 2005).

Kiyonari, T., Shimoma, E., & Yamagishi, T. (2004). Second-order punishment in a one-shot social dilemma. Poster presentation at the28th International Congress of Psychology, Beijing, China (August 2004).

Komorita, S. S., & Parks, C. D. (1995). Interpersonal relations: mixed-motive interaction. *Annual Review of Psychology, 46*, 183-207.

Krupp, D., DeBruine, L., & Barclay, P. (2005). A cue of kinship affects cooperation in a "tragedy of the commons". Talk presented at the 17th Annual Meeting of the Human Behavior & Evolution Society, Austin, Texas (June 2005).

Kurzban, R. and Houser, D. (2001). Individual Differences in Cooperation in a Circular Public Goods Game, European Journal of Personality 15, S37-S52.

Kurzban, R., & Houser, D. (2005). Experiments investigating cooperative types in humans: a complement to evolutionary theory and simulations. *PNAS, 102*, 1803-1807.

Kurzban, R., O'Brien, E., Malmgren-Samuel, E., Tusen, E., & Weissberger, A. (2004). Third-party punishment: punitive sentiment or reputation? Talk presented at the 16th Annual Meeting of the Human Behavior & Evolution Society, Free University of Berlin, Germany (July 2004).

Ledyard, J. O. (1995). Public goods: a survey of experimental research. In *The Handbook of Experimental Economics* (Eds. J. H. Kagel & A. E. Roth), pp. 111-194. Princeton, NJ: Princeton University Press.

Leimar, O., & Hammerstein, P. (2001). Evolution of cooperation through indirect reciprocity. *Proceedings: Biological Sciences, 268*, 745-753.

Liberman, V., Samuels, S. M., &.Ross, L. (2004). The name of the game: predictive power of reputations versus situational labels in determining Prisoner's Dilemma Game moves. *Personality and Social Psychology Bulletin, 30*, 1175-1185.

Lotem, A., Fishman, M. A., & Stone, L. (2002). From reciprocity to unconditional altruism through signaling benefits. *Proceedings: Biological Sciences, 270*, 199-205.

Lotem, A., Wagner, R. H., & Balshine-Earn, S. (1999). The overlooked component of nonsignaling behavior. *Behavioral Ecology, 10*, 209-212.

Masclet, D., Noussair, C., Tucker, S., & Villeval, M.-C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review, 93*, 366-380.

Mashima, R., & Takahashi, N. (2003). The emergence of indirect reciprocity: is the standing strategy the answer? Talk at the 15th Annual Meeting of the Human Behavior and Evolution Society, University of Nebraska, Lincoln, NE (June 2003).

Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Science, 246*, 15-18.

Mazur, A., & Booth, A. (1998). Testosterone and dominance in men. *Behavioral and Brain Sciences, 21*, 353-397.

McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2004). Sustaining cooperation in trust games. Submitted to *Economic Journal*. Available online at: http://www.ices-gmu.org/pdf/materials/415.pdf

McCusker, C., & Carnevale, P. J. (1995). Framing in resource dilemmas: loss aversion and the moderating effects of sanctions. *Organizational Behavior and Human Decision Processes, 61*, 190-201.

McNamara, J. M., Barta, Z., & Houston, A. I. (2004). Variation in behaviour promotes cooperation in the Prisoner's Dilemma game. *Nature, 428*, 745-748.

McNamara, J. M., & Houston, A. I. (2002). Credible threats and promises. *Philosophical Transactions of the Royal Society of London: Series B, 357*, 1607-1616.

Mealey, L. (1985). The relationship between social status and biological success: a case study of the Mormon religious hierarchy. *Ethology and Sociobiology, 6*, 249-257.

Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Evolution and Human Behavior, 17*, 119-128.

Messick, D. M., & Brewer, M. B. (1983). Solving social dilemmas: a review. In *Review of Personality and Social Psychology* (Eds. L. Wheeler & P. Shaver), pp. 11-44. Beverly Hills, CA: Sage Publications.

Milinski, M., Külling, D., & Kettler, R. (1990). Tit for Tat: sticklebacks (*Gasterosteus aculeatus*) "trusting" a cooperating partner. *Behavioral Ecology, 1*, 7-11.

Milinski, M., Pfluger, D., Külling, D., & Kettler, R. (1990). Do sticklebacks cooperate repeatedly in reciprocal pairs? *Behavioral Ecology and Sociobiology, 27*, 17-21.

Milinski, M., Semman, D., Bakker, T. C. M., & Krambeck, H.-J. (2001). Cooperation through indirect reciprocity: image scoring or standing strategy? *Proceedings: Biological Sciences, 268*, 2495-2501.

Milinski, M., Semmann, D., & Krambeck, H.-J. (2002a). Reputation helps solve the "tragedy of the commons". *Nature, 415*, 424-426.

Milinski, M., Semmann, D., & Krambeck, H.-J. (2002b). Donors to charity gain in both indirect reciprocity and political reputation. *Proceedings: Biological Sciences, 269*, 881-883.

Miller, G. (2000). *The Mating Mind: How sexual selection shaped the evolution of human nature.* New York: Doubleday.

Mims, P. R., Hartnett, J. J., & Nay, W. R. (1975). Interpersonal attraction and help volunteering as a function of physical attractiveness. *The Journal of Psychology, 89*, 125-131.

Monteresso, J., Ainslie, G., Toppi Mullen, P. A.-C. P., Gault, B. (2003). The fragility of cooperation: A false feedback study of a sequential iterated prisoner's dilemma. *Journal of Economic Psychology, 23*, 437-448.

Muller, A., & Vickers, M. (1996). Communication in a common pool resource environment with probabilistic destruction. McMaster University, Department of Economics, Working Paper 96-06.

Myers, D. G., & Spencer, S. J. (2001). *Social Psychology (Canadian Edition)*. Toronto, ON: McGraw-Hill Ryerson.

Noë, R. (1990). A Veto game played by baboons: a challenge to the use of the Prisoner's Dilemma as a paradigm for reciprocity and cooperation. *Animal Behaviour*, *39*, 78-90.

Nowak, M. A., Page, K. M., & Sigmund, K. (2000). Fairness versus reason in the Ultimatum Game. *Science*, *289*, 1773-1775.

Nowak, M. A., & Sigmund, K. (1992). Tit for tat in heterogenous populations. *Nature*, *355*, 250-253.

Nowak, M. A., & Sigmund, K. (1993). A strategy of win-stay, lost-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*, *364*, 56-58.

Nowak, M. A., & Sigmund, K. (1998a). Evolution of indirect reciprocity by image scoring. *Nature*, *393*, 573-577.

Nowak, M. A., & Sigmund, K. (1998b). The dynamics of indirect reciprocity. *Journal of Theoretical Biology*, *194*, 561-574.

Oda, R. (2001). Sexually dimorphic mate preference in Japan. *Human Nature*, *12*, 191-206.

Olendorf, R., Getty, T., & Scribner, K. (2004). Cooperative nest defence in red-winged blackbirds: reciprocal altruism, kinship, or by-product mutualism? *Animal Behaviour*, *271*, 177-182.

Oliver, P. (1980). Rewards and punishments as selective incentives for collective action: theoretical investigations. *American Journal of Sociology*, *85*, 1356-1375.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, UK: Cambridge University Press.

Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, *86*, 404-417.

Palmer, C. T. (1991). Kin-selection, reciprocal altruism, and information sharing among Maine lobstermen. *Ethology and Sociobiology, 12*, 221-235.

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.

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

Patton, J. (2004). Coalitional effects on reciprocal fairness in the Ultimatum Game: A case from the Ecuadorian Amazon. In *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies* (Eds. J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, & H. Gintis), pp 96-124. Oxford, UK: Oxford University Press.

Patton, J. Q. (2005). Meat sharing for coalitional support. *Evolution and Human Behavior, 26*, 137-157.

Peterson, L., Ridley-Johnson, R., & Carter, C. (1984). The supersuit: An example of structured naturalistic observation of children's altruism. *The Journal of General Psychology, 110*, 235-241.

Petrinovich, L., O'Neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions: toward and evolutionary ethics. *Journal of Personality and Social Psychology, 64*, 467-478.

Price, M. E. (2003). Pro-community altruism and social status in a Shuar village. *Human Nature, 14*, 191-208.

Price, M. E. (2005). Punitive sentiment among the Shuar and in industrialized societies: cross-cultural similarities. *Evolution and Human Behavior, 26*, 279-287.

Rabbie, J.M., Schot, J.C., & Visser, L. (1989). Social identity theory: A conceptual and empirical critique from the perspective of a behavioural interaction model. *European Journal of Social Psychology, 19*, 171-202.

Rege, M., & Telle, K. (2004). The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics, 88*, 1625-1644.

Roberts, G. (1998). Competitive altruism: from reciprocity to the handicap principle. *Proceedings: Biological Sciences, 265*, 427-431.

Roberts, G., & Renwick, J. S. (2003). The development of cooperative relationships: an experiment. *Proceedings: Biological Sciences, 270*, 2279-2283.

Roberts, G., & Sherratt, T. N. (1998). Development of cooperative relationships through increasing investment. *Nature, 394*, 175-179.

Rohner, R. P., & Rohner, E. C. (1970). *The Kwakiutl: Indians of British Columbia*. New York, NY: Holt, Rinehart and Winston.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Roth, A. E. (1995). Bargaining experiments. In *The Handbook of Experimental Economics* (Eds. J. H. Kagel & A. E. Roth), pp. 253-348. Princeton, NJ: Princeton University Press.

Rutherford, M. D. (*in press*). The effect of social role on Theory of Mind reasoning. *British Journal of Psychology*.

Schino, G. 2001 Grooming, competition and social rank among female primates: a meta-analysis. *Animal Behaviour 62*, 265-271.

Schultheiss, O. C., Campbell, K. L., & McClelland, D. C. (1999). Implicit power motivation moderates men's testosterone responses to imagined and real dominance success. *Hormones and Behavior, 36*, 234-241.

Sefton, M., Shupp, R., & Walker, J. (2002). The effects of rewards and sanctions in provision of public goods. *Working Paper W00-16*, Centre for Decision Research and Experimental Economics, Nottingham, U.K.

Seinen, I., & Schram, A. (*forthcoming*). Social status and group norms: indirect reciprocity in a helping experiment. *European Economic Review*.

Semmann, D., Krambeck, H.-J., & Milinski, M. (2004). Strategic investment in reputation. *Behavioral Ecology and Sociobiology, 56*, 248-252.

Seyfarth, R. M. 1977 A model of social grooming among adult female monkeys. *Journal of Theoretical Biology 65*, 671-698.

Sheldon, K. M., Skaggs Sheldon, M., & Osbaldiston, R. (2000). Prosocial values and group assortation. *Human Nature, 11*, 387-404.

Sherratt, T. N., & Roberts, G. (1999). The evolution of quantitatively responsive cooperative trade. *Journal of Theoretical Biology, 200*, 419-426.

Sigmund, K., Hauert, C., & Nowak, M. A. (2001). Reward and punishment. *PNAS, 98*, 10757-10762.

Simpson, J.A., & Gangestad, S.W. (1991). Individual differences in sociosexuality: Evidence for convergent and discriminant validity. *Journal of Personality and Social Psychology, 60*, 870-883.

Smeesters, D., Warlop, L., Van Avermaet, E., Corneille, O., & Yzerbyt, V. (2003). Do not prime hawks with doves: the interplay of construct activation and consistency of Social Value Orientation on cooperative behavior. *Journal of Personality and Social Psychology Bulletin, 84*, 972-987.

Smith, E. A. (2003). Human cooperation: perspectives from behavioral ecology. In *Genetic and Cultural Evolution of Cooperation* (Ed. P. Hammerstein), pp. 401-428. Cambridge, MA: MIT Press.

Smith, E. A. (2004). Why do good hunters have higher reproductive success? *Human Nature, 15*, 343-364.

Smith, E. A., & Bliege Bird, R. (2000). Turtle hunting and tombstone opening: public generosity as costly signaling. *Evolution and Human Behavior, 21*, 245-262.

Smith, E. A., & Bliege Bird, R. (*in press*). Costly signaling and cooperative behavior. To appear in *The Moral Sentiments: Theory, Evidence, and Policy* (Eds. S. Bowles, R. Boyd, E. Fehr, & H. Gintis). Cambridge, MA: MIT Press.

Smith, E. A., Bliege Bird, R., & Bird, D. W. (2003). The benefits of costly signaling: Meriam turtle hunters. *Behavioral Ecology, 14*, 116-126.

Snodgrass, S. E. (1985). Women's intuition: The effect of subordinate role on interpersonal sensitivity. *Journal of Personality and Social Psychology, 49*, 146-155.

Snodgrass, S. E. (1992). Further effects of role versus gender on interpersonal sensitivity. *Journal of Personality and Social Psychology, 62*, 154-158.

Sober, E., & Wilson, D. S. (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Sosis, R. (2000). Costly signaling and torch fishing on Ifaluk atoll. *Evolution and Human Behavior, 21*, 223-244.

Sosis, R., & Alcorta, C. (2003). Signaling, solidarity, and the sacred: the evolution of religious behavior. *Evolutionary Anthropology, 12*, 264-274.

Staller, A., Sloman, S. A., & Ben-Zeev, T. (2000). Perspective effects in nondeontic versions of the Wason selection task. *Memory and Cognition, 28*, 396-405.

Stone, V. E., Cosmides, L., Tooby, J., Kroll, N., & Knight, R. T. (2002). Selection impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *PNAS, 99*, 11531-11536.

Strassberg, D. S., & Holty, S. H. (2003). An experimental study of women's internet personal ads. *Archives of Sexual Behavior, 32*, 253-260.

Strohmetz, D. B., Rind, B., Fisher, R., & Lynn, M. (2002). Sweetening the till: the use of candy to increase restaurant tipping. *Journal of Applied Social Psychology, 32*, 300-309.

Tajfel, H., Billig, M.G., Bundy, R.P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*, 149-178.

Takahashi, N., & Mashima, R. (2004). The effect of perception errors on the emergence of generalized exchange. Talk at the 16th Annual Meeting of the Human Behavior and Evolution Society, Free University of Berlin, Berlin (July 2004).

Tessman, I. (1995). Human altruism as a courtship display. *Oikos, 74*, 157-158.

Thiessen, D., Young, R. K., & Burroughs, R. (1993). Lonely hearts advertisements reflect sexually dimorphic mating strategies. *Ethology and Sociobiology, 14*, 209-229.

Tinbergen, N. (1968). On war and peace in animals and man. *Science, 160*, 1411-1418.

Tooby, J., & Cosmides, L. (1996). Friendship and the Banker's Paradox: other pathways to the evolution of adaptations for altruism. *Proceedings of the British Academy, 88*, 119-143.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences, 16*, 495-552.

Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*, 35-57.

Urbaniak, G.C., & Kilmann, P.R. (2003). Physical attractiveness and the "Nice Guy Paradox": Do nice guys really finish last? *Sex Roles, 49*, 413-426.

Utz, S., Ouwerkerk, J. W., & Van Lange, P. A. M. (2004). What is smart in a social dilemma? Differential effects of priming competence on cooperation. *European Journal of Social Psychology, 34*, 317-332.

Van Lange, P.A.M., Ouwerkerk, J.W., & Tazelaar, J.A. (2002). How to overcome the detrimental effects of noise in social interaction: The benefits of generosity. *J. Pers. & Soc. Psych. 82*, 768-780.

Van Soest, D.P. and Vyrastekova, J. (2004). Economic ties and social dilemmas: an economic experiment. *CentER Discussion Paper 2004-55*, CentER, University of Tilburg, Netherlands.

Van Vugt, M., & De Cremer, D. (1999). Leadership in social dilemmas: The effects of group identification on collective actions to provide public goods. *Journal of Personality and Social Psychology, 76*, 587-599.

Watts, D. P. (2002). Reciprocity and interchange in the social relationships of wild male chimpanzees. *Behaviour, 139*, 343-370.

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.

Wiederman, M. (1993). Evolved gender differences in mate preferences: Evidence from personal advertisements. *Ethology and Sociobiology, 14*, 331-352.

Wiessner, P. (2003). The perils and pleasures of punishment among foragers. Talk presented at the 15th Annual Meeting of the Human Behavior & Evolution Society, University of Nebraska, Nebraska (June 2003).

Wilkinson, G. S. (1984). Reciprocal food sharing in the vampire bat. *Nature, 308*, 181-184.

Williams, G. C. (1966). *Adaptation and Natural Selection*. Princeton, NJ: Princeton University Press.

Wilson, D. S. (1998). Hunting, sharing, and multilevel selection: the tolerated-theft model revisited. *Current Anthropology, 39*, 73-97.

Wilson, D. S. (2004). What is wrong with absolute fitness? *Trends in Ecology and Evolution, 19*, 245-248.

Wood, B., & Hill, K. (2000). A test of the "showing-off" hypothesis with Ache hunters. *Current Anthropology, 41*, 124-125.

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology, 51*, 110-116.

Yamagishi, T. (2003). The group heuristic: A psychological mechanism that creates a self-sustaining system of generalized exchanges. Paper prepared for workshop on "The Co-evolution of Institutes and Behavior", Sante Fe Institute, Jan. 10-12, 2003.

Yamagishi, T., Foddy, M., Makimura, Y., Matsuda, M., Kiyonari, T., Platow, M. J. (*in press*). Comparisons of Australians and Japanese on group-based cooperation. *Asian Journal of Social Psychology, 8*.

Yamagishi, T., & Kiyonari, T. (2000) The group as the container of generalized reciprocity. *Social Psychology Quarterly, 63(2)*, 116-132.

Zahavi, A. (1975). Mate selection – A selection for handicap. *Journal of Theoretical Biology, 53*, 205-214.

Zahavi, A. (1977a). The cost of honesty (further remarks on the handicap principle). *Journal of Theoretical Biology, 67*, 603-605.

Zahavi, A. (1977b). Reliability in communication systems and the evolution of altruism. In *Evolutionary Ecology* (Eds. B Stonehouse & C. Perrins), pp. 253-259. Baltimore, MD: University Park Press.

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