# REPUTATION AND THE EVOLUTION OF GENEROUS BEHAVIOR

# REPUTATION AND THE EVOLUTION OF GENEROUS BEHAVIOR

**PAT BARCLAY**
**EDITOR**

**Nova Science Publishers, Inc.**
*New York*

**LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA**

ISBN: 978-1-61668-153-1

Published by Nova Science Publishers, Inc.

; New York

# CONTENTS

# PREFACE

Why are humans sometimes generous to each other? This question remains a central question in social psychology, evolutionary biology, experimental economics, sociology, and other disciplines: if generous individuals provide benefits to others at a cost to themselves, then these costs imply that generosity will not evolve, not be learned, and/or not be chosen rationally. Despite these costs, humans often do help others at a cost to themselves. In fact, humans are perhaps the most cooperative of all species when it comes to generosity towards non-relatives.

Many theories have been proposed to account for the existence of generous sentiment in nature, most of which predict that such sentiments will only exist if possessing them and acting on them brings benefits (i.e. rewards during development and/or on an evolutionary timescale). *Reputation and the Evolution of Generous Behaviour* discusses recent theoretical and empirical advances that attempt to explain the existence of generous sentiment: why do humans possess a psychology that causes them to occasionally help others at a cost to themselves, and even help groups of non-relatives? Barclay argues that benefits to one's reputation can be a major factor causing the existence of generous sentiment. In support of this argument, he brings together theory and empirical research from evolutionary biology and evolutionary psychology, experimental economics, social psychology, anthropology, and game theory. In the process of discussing reputation, he dissects the proposition that human generosity must have involved natural selection acting on entire groups (a process known as "group selection").

Reputation can support generous sentiment in a number of ways. Group generosity may be supported by systems of indirect reciprocity, where people help those who help others, and those who do not help helpers are excluded

from this circle of benefits. Cooperators may also benefit from preferentially interacting with other cooperators. Theories of costly signaling suggest that generous acts may function (with or without intention) as signals of unobservable qualities such as resources or cooperative intent, such that generous individuals may benefit (possibly unintentionally) from the advertisement of such qualities – these signals will be honest so long as the cost of the generosity is sufficiently high to discourage such behaviour in individuals who do not actually possess such qualities. This even leads to the prediction that individuals will compete to be more generous than others in order to increase their desirability as partners, and hence their position in the marketplace for social partners. The empirical evidence suggests that generous individuals can benefit in many ways for their acts, such that these reputational benefits can help explain why cooperative sentiment evolved and/or is learned through social reinforcement. By examining the forces that sustain generosity, we can use this information to promote situations that foster generosity.

*Chapter 1*

# 1. INTRODUCTION

## 1.1. GENERAL INTRODUCTION

Why are humans sometimes generous to each other? This is a very old question, and remains a central question in social psychology, evolutionary biology, experimental economics, sociology, and other disciplines. We can rephrase this question as: why would an organism ever do something that is costly to itself yet beneficial to another? Evolutionary biologists have long been puzzled by the existence of behaviour that appears to be costly to individuals, so as such they have sought to explain the existence of generous behaviour by considering how natural selection might result in the evolution of a psychology that causes generosity.

In this book, I continue that tradition by reviewing some possible benefits that generous individuals might receive for acting generously towards non-relatives, in order to infer how the sentiments that cause generous behaviour could have evolved and/or be learned by individual or cultural learning (Lehmann et al., 2008). An understanding of why generosity exists at all can help researchers to discover the factors that maintain it, situations where it occurs, and how to promote generosity and other forms of cooperation.

## 1.2. AVOIDING CONFUSION AND MISINTERPRETATION

I would love to be able to dive right into the causes of generosity, but I feel that some clarifications are necessary first. From past experience, many people misunderstand evolutionary arguments about altruism and generosity. This is unfortunate, because evolutionary theory has a lot to add to our

understanding about generosity. However, this misunderstanding is somewhat predictable, given that evolutionists use different terms and ask different types of questions that other researchers often don't ask. As such, some of this misunderstanding is due to disciplinary differences in the definitions of "altruism" and "generosity", and some is due to confusion between psychological motivations and evolutionary function. These need to be clarified before actually discussing generosity and how generous sentiment could evolve.

### 1.2.1. What Phenomena am I Talking about? How am I Defining Generosity?

The word "altruism" is used very differently across fields. Some fields define altruism in terms of the sentiments underlying the behavior and the intentions of the actor, whereas other fields focus only on the effect of the behaviour on the actor and the recipient. Even when focusing only on the effects of behaviour, some researchers refer only to the immediate costs and benefits while other researchers refer to lifetime costs and benefits. Furthermore, some researchers define the costs and benefits as being relative to the broader population whereas other researchers define the costs of "altruism" as being relative to one's local group (regardless of the effects relative to the broader population). These different definitions have caused much confusion, and there has been much disagreement simply because people have been arguing about different phenomena that happen to have the same name (for a discussion of this, see West et al., 2007; or Bshary and Bergmüller, 2008). I will attempt to avoid this confusion by limiting my use of the word "altruism", and instead use the word "generosity" as a behavioural phenomenon; although this term does have some connotations, it has far less baggage than the word "altruism".

Rather than engage in lengthy and possibly fruitless debate over what the correct definition of "generosity" is or should be, I will simply describe how I will use the word and what phenomena I am defining as "generous". In this book, I am using a definition of generosity that looks only at the immediate costs to the giver (relative to the overall population) and the benefits to the recipient, rather than the particular motivations that underlie such behaviour. By this behavioural definition, generosity does not depend on what psychological factors motivate any helping or giving, nor does it depend upon an absence of long-term benefits to the giver. Thus, this definition includes

people giving in order to receive later or in order to be seen as good. If readers disagree with this usage and disagree with calling such acts "generous", then I will simply invite them to substitute their own terms for the phenomena that I am discussing, whether that term be "prosociality", "cooperation", "other-benefiting behaviour", "giving", or simply "helping others". This way, the focus will be on the behavioural phenomenon rather than on specific terms.

### 1.2.2 Different Q's: Psychological Mechanisms, Development, and Evolutionary Function

While differing definitions of "altruism" and "generosity" or even "cooperation" have caused many misunderstandings, a much bigger cause of misunderstanding is that evolutionary biologists and evolutionary psychologists often ask different types of questions than do researchers in other fields. For example, many researchers investigate the psychological motivations that cause apparently generous behaviour and the developmental forces that forge those motivations, or the situational factors that trigger them. These are indeed important areas of research. However, there is another type of question that can be asked, namely that of *function*: why do such motivations exist (if indeed they do)? Why isn't the brain set up in a different way that always causes selfishness? If for example, people feel good about helping, why do they feel good and why aren't humans built to feel terrible about incurring costs on themselves in order to benefit others? Evolutionists often seek to answer this type of question.

I will spend much time discussing the costs and benefits of behaving generously, and especially the benefits caused by enhanced reputation. By investigating the benefits that individuals receive for behaving generously, this does not mean to imply that generous acts are consciously deliberated, nor does it imply that people always have ulterior motives and seek to benefit from being nice to others. People may be genuinely 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 an altruistic manner. Also, if a generous act does happen to bring benefits to the giver at a later point in time, this does not mean the act was not generous at the particular time it was performed. Rather than investigating the particular motivational mechanisms that underlie such behaviour, this book covers why people might have the sentiments that cause such behaviour, and investigates the cues and incentives that trigger generosity in order to make inferences

about what selective forces might have shaped the capacity to develop such motivations. In other words, the question becomes: what causes cooperative and generous sentiment to arise and be maintained despite the costs of generosity? Why does such sentiment exist at all? Why aren't people always as selfish as possible?

This question of *evolutionary function* is separate and complementary to the questions of what the particular *psychological motivations* or *developmental causes* are (Tinbergen, 1968). According to this distinction, the *psychological motivations* are what goes on in the head of any particular person that might cause him/her to behave generously, the *developmental causes* are the paths in development that cause those psychological motivations to come to be (e.g. how much influence does learning have), and the *function* is the reason that those psychological motivations and developmental paths exist at all and why they are maintained despite the costs of generous behaviour. I will focus mostly on the latter.

It is entirely possible that (some) people possess entirely unselfish motives. However, even if they do, that would not necessarily say anything about the evolution of such motives (Sober and Wilson, 1998). If having cooperative sentiments and acting on them tends to bring benefits to people, then such sentiments will have tended to increase in prevalence in populations via biological evolution across generations and/or learning (e.g. Frank, 1988).

Although I am primarily discussing the evolution of generous sentiment, this does not necessarily imply that people do not learn how much generosity to perform (Lehmann et al., 2008): in fact, 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 generous acts and the benefit that they bring. In this situation, learning is the *developmental cause*, but we still need to investigate the *function* by asking why people learn to be generous instead of being selfish – why doesn't the cost of generosity cause people to learn to be selfish? Are there rewards or reinforcements that cause people to learn generosity? If so, why do such reinforcements occur? It is important to investigate the types of benefits that altruists might receive, in order to determine why humans possess a psychology that causes them to be generous to others.

*Chapter 2*

# WHY IS GENEROSITY PUZZLING?

For decades, evolutionary biologists have sought to explain the existence of apparently generous behaviour in nature. When an organism acts generously, it benefits others at a cost to itself. As long as such benefits and costs translate in some way to gains and losses in reproduction (however small those gains and losses), then generous types would be at a selective disadvantage relative to selfish types. Generous organisms would leave fewer offspring than selfish types, causing a decrease in the proportion of organisms that possess causal mechanisms for such behaviour. A similar thing could happen within the lifespan of a given organism via learning by conditioning – if generosity is costly but brings no benefits then one might expect this behavior to be extinguished. 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 generous types would have higher fitness (i.e. leave more descendants) than the generous types. 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 and 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 and Wilson, 1983; Dugatkin, 2004), and several have focused on humans (e.g. Betzig and Turke, 1986; Daly and Wilson, 1988; DeBruine, 2002; Grayson, 1993; Hames, 1987; Krupp et al., 2008; Petrinovich, O-Neill, and Jorgensen, 1993; Stewart-Williams, 2007).

Inclusive fitness theory is extraordinarily powerful, and is one of the cornerstones in studying the evolution of social behaviour. As powerful as this idea is, it is unlikely to explain all generous behaviour because many such acts appear to be systematically directed towards non-kin. The cost of such generosity 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.

*Chapter 3*

# DYADIC RELATIONSHIPS AND "RECIPROCAL ALTRUISM"

## 3.1. DIRECT RECIPROCITY

Trivers (1971) introduced the concept of "reciprocal altruism", in which individuals who reciprocate generous acts towards each other can outcompete others who do not. This process requires that generous types can distinguish between others and direct their generosity towards others that have reciprocated in the past. Organisms that are inclined towards reciprocity will 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 reciprocity 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 payoff-maximizing strategy in any given round, 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 reciprocity can provide a selective advantage.

Much work has since used the PD to model cooperative interactions. For example, the presence of occasional defectors due to mutation or error allows conditional cooperators (such as Tit for Tat or others using reciprocity) to dominate unconditional cooperators (McNamara, Barta, and Houston, 2004; Nowak and Sigmund, 1992). Tit for Tat itself can be dominated by strategies that are more forgiving or that will exploit unconditional cooperators (Nowak and Sigmund, 1992, 1993). Tit for Tat and other conditional cooperators need to achieve some critical frequency in order to invade a population. Below this critical frequency, they will not encounter each other sufficiently often to overcome the disadvantage of being "suckered" on their first interaction with each defector (Dawkins, 1976). This critical threshold is easier to achieve if there is some assortment among cooperators, for example because of kinship (Dawkins, 1976), or if there already exists in the population some "suspicious" conditional cooperators who do not cooperate until their partner does (e.g. "Suspicious Tit for Tat", Boyd and Lorberbaum, 1987).

Most work on direct reciprocity has used a PD-like situation where players can only do one of two discrete moves (cooperate or defect). In the real world, organisms usually have more than these two discrete options available. 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 and Sherratt, 1998; Sherratt and Roberts, 1999). This strategy prevents individuals from being exploited too much in early rounds by non-cooperators, yet still allows high levels of cooperation to be achieved over time. In experimental games, humans do seem to follow this strategy of "raising the stakes" (Roberts and Renwick, 2003), especially with strangers with whom they have not interacted (Majolo et al., 2006). This strategy is robust against subtle cheaters who slightly undercut levels of cooperation, even when there is some ambiguity over partners' exact contributions (Van den Bergh and Dewitte, 2006).

Many researchers have claimed to find evidence of direct reciprocity 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, and Kettler, 1990; Milinski, Pfluger, Külling, and 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, and Hill, 2000; Hauser, Chen, Chen,

and 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, and 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 reciprocity in non-human animals or at least non-primates (Hammerstein, 2003; Noë, 1990), possibly because of a lack of the necessary cognitive requirements (Stevens and Hauser, 2004), so it is currently fair to say that much of the evidence for reciprocity in non-humans is equivocal, or at least that reciprocity is not as widespread as many researchers would like to believe. Nevertheless, recent work in birds appears to be free from experimental confounds and demonstrates that some bird species may provide less help towards those who have refused to help them (Krams et al., 2008; Wheatcroft and Price, 2008).

The evidence for reciprocity 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, and McCabe, 1995; Cox, 2004; Fehr, Fischbacher and Gächter, 2002; Komorita and Parks, 1995; Roberts and 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 and Minnegal, 1997; Gurven, Hill, Kaplan, Hurtado, and Lyles, 2000; Gurven, Allen-Arave, Hill, and Hurtado, 2001; Patton, 2005), labour exchange (Hames, 1987), restaurant tipping (Strohmetz, Rind, Fisher, and 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. In one well-known study, Mealey, Daood, and Krage (1996) found that people had better memory for the faces of putative low-status cheaters than for other people, although other authors have had difficulty replicating this alleged enhanced memory for the faces of cheaters (Barclay and Lalumière, 2006; Mehl and Buchner, 2008; see Barclay, 2008 for a review). 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, 2008; Cheng and Holyoak, 1989; Fodor, 2000; Staller, Sloman, and 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. Barclay, 2008; Fischbacher, Gächter, and Fehr, 2001; Monteresso, Ainslie, Toppi Mullen, and Gault, 2003). Thus, humans may possess cognitive mechanisms that function to support reciprocity even if those mechanisms also allow humans to solve other problems or evolved for somewhat more general purposes such as general reputation tracking.

## 3.2. INDIRECT RECIPROCITY

### 3.2.1. Evidence for Indirect Reciprocity

Direct reciprocity occurs when recipients of generosity can (and do) reciprocate generous acts directly to the giver. Sometimes generosity can be reciprocated indirectly, i.e. by individuals other than the beneficiary of the help (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. People may use some combination of personal experience and observations of others helping and non-helping in order to make their decision about whether to cooperate with someone (Roberts, 2008).

In an early mathematical model of indirect reciprocity, Nowak and Sigmund (1998a, b) used a situation where helping someone else increases one's "image score" (i.e. reputation) and non-helping decreases one's image score. Agents in this model only cooperated with others whose score is above a threshold "image score", where the threshold could evolve to range from a very low threshold (unconditional cooperation) to very high (unconditional defection). Nowak and Sigmund showed that some degree of generosity would evolve in such circumstance and that such an "image scoring" 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. These findings suggest that humans do engage in some sort of indirect reciprocity, and other researchers using similar methods have reported similar results (Bolton, Katok, and Ockenfels, 2005; Seinen and Schram, 2006).

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 help them to ensure their continued presence in a group (Kaplan and Hill, 1985; Tooby and Cosmides, 1996).

In addition to humans, there is also evidence for indirect reciprocity in the interactions between cleaner fish and the client fish that they clean (Bshary and Grutter, 2006). Client fish spend more time next to cleaners who are observed to cooperate with past clients (i.e. feed on clients' ectoparasites rather than on client mucus, which is feeding against their preference). The cleaner fish themselves are more cooperative when observed than when unobserved (i.e. they spend more time feeding on ectoparasites rather than client mucus when observed).

### 3.2.2. What Form of Indirect Reciprocity?

Although Nowak and Sigmund's (1998a,b) "image scoring" models of indirect reciprocity are easy to understand, other theorists have criticized these particular "image scoring" models on the grounds that they can be invaded by individuals who are generous to anyone (cooperator or defector) whenever they themselves need to improve their reputation. These authors have presented alternative models of indirect reciprocity (e.g. Leimar and Hammerstein, 2001; Mashima and Takahashi, 2003; Panchanathan and Boyd, 2003; Takahashi and Mashima, 2004, for a review see Nowak and Sigmund, 2005), that they claim 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 acquire "bad standing" by defecting on cooperators, but they 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. Such "standing strategies" are evolutionarily stable even when information about partners is available directly from past experience, whereas the presence of this past experience undermines "image scoring" models of indirect reciprocity (Roberts, 2008). In exhaustive searches of all possible ways of playing simple indirect reciprocity games and assigning "goodness" to other players, Ohtsuki and Iwasa (2004, 2006, 2007) have discovered what they call the "leading eight" – eight strategies that outperform all others. All strategies in the leading eight start out cooperative, assume that others have good reputation until proven otherwise, give to cooperators, retaliate against defectors, forgive defections against defectors (i.e. they are variants of "standing strategies"), and forgive defectors who start being cooperative again.

There have been attempts to determine which of "image scoring" or "standing strategies" provide a more accurate description of what people actually do when engaging in indirect reciprocity (Bolton et al., 2005; Milinski, Semmann, Bakker, and Krambeck, 2001). In a donations game where participants had information on their potential recipients' donation histories and the history of their recipients' recipients' donations, Milinski et al. (2001) found that people tended to treat defections against non-cooperators the same as defections against cooperators. Other unpublished studies by Milinski and colleagues have similarly failed to find evidence that humans use standing strategies for indirect reciprocity (Manfred Milinski, personal communication, Aug. 30th, 2004). Bolton et al. (2005) gave people information on what their potential recipients had done to whom in the past, and found that participants donated more often when they had information about who their recipients had given or refused help to (i.e. whether they had given or refused to give to cooperators or defectors). However, this could simply be because this second-order information gives participants a better estimate of the proportion of cooperators in the population: if one knows that a person gave to a cooperator, then one knows that there are at least two other cooperators in the population. This is especially important when participants only know their

partner's actions in one previous round instead of many rounds, because it gives a better estimate of the general level of cooperation. This knowledge of the number of cooperators can easily raise donation levels, given that people tend to cooperate if they think that others will do likewise (Dawes, McTavish, and Shaklee, 1977; Smeesters et al., 2003). Bolton et al. (2005) found that giving to cooperators increases one's chances of receiving donations, and defecting upon cooperators decreases one's chances of receiving donations. Defecting on defectors had a non-significant negative effect, and they did not report whether defecting on defectors had a significantly different effect on one's likelihood of receiving a donation than defecting on cooperators did. Thus, it is ambiguous whether participants treated such acts like regular defections (as image scoring would predict) or like justifiable defections and therefore the same as cooperating (as standing strategies would predict).

Thus, the available studies testing between image scoring and standing strategies provide mixed evidence as to which model is a more accurate description of human behaviour, despite the intuitive appeal of standing strategies and the models supporting their existence. Perhaps studies that use a strong moral framing might provide evidence that humans do something akin to standing strategies and distinguish between justified and unjustified defection, because in such studies it would be much clearer that non-cooperators "deserve" to be defected upon. For example, instead of choosing between "cooperate" and "defect" or "give and "not give", participants could choose between playing "patriot" and "traitor" (or "terrorist") or something similar.

In a related study on people's willingness to pay to punish others by directly imposing costs on them, Barclay (2006) found that people distinguish between "justified" punishment of defectors and "unjustified" punishment of non-defectors, but only if they had enough experience with defectors in the task such that they understood that punishment of defectors benefited others. A similar thing could be occurring in indirect reciprocity – the motives for non-cooperation can be unclear, such that it must be clear that defecting on defectors is intended as a punishment rather than as a selfish act. 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 humans do follow some form of indirect reciprocity.

*Chapter 4*

# THE PROBLEM OF COLLECTIVE ACTION

## 4.1. INTRODUCTION TO PUBLIC GOODS

Systems of direct and indirect reciprocity both rely on individuals being able to target their generosity 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 and Holt, 1993, Messick and 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. Many students will be familiar with the public goods of clean kitchens in residences or work on university group projects: everyone would prefer clean kitchens to dirty ones and would prefer that *someone* worked on the group project, but might also prefer that *someone else* did the work and the cleaning, such that the dirt builds up and the work gets procrastinated. Scientists will recognize peer review as a public good: good science depends on thoughtful and timely peer reviews, but they take time, which some are unwilling to provide (Hauser and Fehr, 2007).

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 radio, national defense, scientific research), ancestral humans would have faced many potential public goods situations such as group defense and vigilance 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, and Bird, 2001; Hawkes, 1991, 1993; Hill and 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, and Kushnik; 2002; Bliege Bird and Smith, 2005; Hawkes, O'Connell, and Blurton Jones, 2001a, b; Hill and Kaplan, 1988; Kaplan and Hill, 1985), which seems to preclude reciprocity. 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 people 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 and Messick, 2000). Participants usually contribute between 40% and 60% of their

endowments in such games, and contributions typically drop with repeated play (Davis and 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, and Fehr, 2004). Theorists and researchers in evolutionary biology, social psychology, political science, sociology, and economics are all interested in the factors that promote cooperation and prevent the drop in contributions.

## 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 and Gächter, 2000, 2002; Ostrom, Walker and 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, and Villeval, 2003). In field settings, low contributors tend to inspire more disapproval and receive more criticism than high contributors (Barr, 2001; Cordell and McKean, 1992; Price, 2005), although very high contributors do sometimes receive punishment (Barr, 2001; see Herrmann et al., 2008). 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 and Gächter, 2002). The costs of punishment may decrease group earnings in the short-term, but can increase earnings in the long-term by preventing the decline of cooperation (Gächter et al., 2008).

Provisioning of public goods can also evolve if contributors are rewarded for their cooperation (Sigmund, Hauert, and Nowak, 2001). Milinski, Semmann and Krambeck (2002a; Semman, Krambeck, and 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 and 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 – the rewarders could even be in a different group entirely (Semmann et al., 2005). These results all clearly show that people will sometimes voluntarily reward those who help provide public goods, and Vyrastekova and van Soest (2008) demonstrated that such rewards are effective at maintaining cooperation in experimental games so long as rewards provide more benefit to the person being rewarded than they cost the person doing the rewarding.

What form might such rewards take? Some rewards may take the form of aid in times of need, and status is another potential reward for generosity. Recipients may pay particular attention to generous individuals such that those individuals 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 generosity 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 and Gil-White, 2001), receive better offers in bargaining and sharing experiments and in simulated markets (Ball and Eckel, 1996, 1998; Ball, Eckel, Grossman, and Zame, 2001; Commins and Lockwood, 1979), have greater control of resources (Betzig, 1988; Ellis, 1993), and are more likely to survive population crashes (Boone and 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 generous people 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 and Rohner, 1970; but see Drucker and 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. Experimental work shows that people who contribute more in experimental cooperative games are rated higher than low contributors on scales measuring status (Hardy and Van Vugt, 2006; Willer, 2009).

## 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 and 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 – in particular, opportunities to retaliate can undermine the existence of punishment (Janssen and Bushman, 2008), and much experimental evidence demonstrates that people retaliate

against received punishment when they are given the opportunity (Barclay, 2006; Cinyabuguma et al., 2006; Denant-Boemont et al., 2007).

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 (e.g. Oliver, 1980; Yamagishi, 1986). 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 and Ockenfels, 2000; Roth, 1995), are sensitive to people taking benefits without paying the appropriate costs (Cosmides and Tooby, 1992), and can learn by observation (Tomasello, Kruger and 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 (Boyd and Richerson, 1992; Henrich and Boyd, 2001), also known as "metanorms" (Axelrod, 1986). 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, and Richerson, 2003; Henrich and 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 and 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 recent studies have instead found a conspicuous lack of second-order punishment (Barclay, 2006; Kiyonari and Barclay, 2008; Kiyonari, Shimoma, and Yamagishi, 2004).

*Chapter 5*

# DOES GROUP SELECTION SOLVE SECOND-ORDER FREE-RIDING?

## 5.1. MODELS OF GROUP SELECTION

There has been much debate recently over the role of "group selection" in the evolution of generous and punitive sentiment. Group selection is when groups of generous individuals do better than groups of selfish individuals, such that generosity supposedly spreads despite the within-group disadvantage for being generous. These group selectionist accounts are often pitted against reputational accounts for the evolution of generous behaviour. If the fitness cost of providing higher-order cooperation (e.g. second-order punishment and rewards) is relatively small, then other selection pressures do not have to be very strong to overcome the fitness disadvantage of providing incentives. As such, relatively minor pressures favouring such behaviours could result in overall selection for the sentiments that promote generosity and the provision of incentives.

Henrich and Boyd (2001) provide one example of such a minor pressure, and note that humans tend to conform to the most common behaviours in their groups. They 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 (Gürerk et al., 2006). Indeed, a computer simulation by Boyd and colleagues (2003) showed

that “altruistic punishment” could evolve via differential survival of groups even though it is individually costly.

The phrase “group selection” sounds appealing, and some authors have attempted to make a big deal of it and claim that it can solve the problem of group cooperation and the problem of second-order free-riding. However, there is much confusion surrounding this issue. Part of this confusion is because there are multiple models that use the words “group selection” in different ways. This confusion is fueled by new definitions of words such as “altruism” as well as imprecision over the specific processes that cause generous strategies to increase in frequency (for a discussion, see West et al., 2007) such that readers may not realize that “group selection” relies on inclusive fitness maximization – in fact it often relies on direct benefits accruing to those who are being “altruistic”. Additional confusion is created by possible confusions between the psychological mechanisms that cause behavior and their evolutionary outcomes (see section 5.2). As such, the usefulness of the term “group selection” has been criticized by multiple authors. I present three different uses as follows.

### 5.1.1. Decreasing Relative Fitness but Increasing Absolute Fitness

In the first class of models of group selection, individuals perform actions that may *increase* their fitness *relative to the global population* but benefit other group members *more*, such that the actors *decrease* their fitness *relative to their local group*. Proponents of these models term such behaviors “altruistic” because of this decrease in *local* relative fitness (ignoring the increase in *global* fitness). For example, individuals may provide a public good that they benefit from but non-providers (“scroungers”) benefit more from, or individuals may increase their group’s cooperation by paying to enforce cooperation with sanctions and incentives, such that they benefit from the higher cooperation but not as much as those who avoid the cost of punishing. This can occur with public good provision, punishment, or even when individuals cooperate to help their groups (and thus themselves) survive through rough times, given that people often have some stake or vested interest in the existence of their groups and in preventing group extinction.

Whenever cooperators do something that decreases their fitness relative to the *local* group but increases their fitness relative to the *global* population, these cooperators are directly benefiting themselves such that they simply

reproduce faster non-cooperators in less cooperative groups. This causes an overall increase in the frequency of cooperators in the population despite the fact that they are doing worse *within each group* (Sober and Wilson, 1998; Wilson, 1998, 2004). This will result in a stable equilibrium with both providers and scroungers that is maintained by the opposing forces of individual and group selection. These models explicitly rely on high levels of gene flow or other exchange between groups so that generosity spreads between populations faster than it is selected against within-populations. In other words, they rely on competition being more "global" than "local" – i.e., individuals must compete to some extent against the global population rather than solely competing with their local population (West et al., 2006).

The reason that proponents call this "group selection" is because the generosity causes the group to do better yet causes non-helpers to do better than helpers *within that group*, even though both do better than non-helpers *in other groups* or better *than they themselves would have been had they not helped*. However, if an individual does something that increases his/her fitness relative to the global population, it is severely misleading to call this "group selection" because the "group" is often not being selected – instead, it is selection of individuals that sometimes happen to be in groups (for a similar point, see West et al. 2007). For example, when a hunter brings in food, he/she may be better off than non-hunters *in other groups* despite doing worse than non-hunters *within his/her group*. The presence of groups – or any benefits to fellow group members – is simply not necessary in these models because the individual benefits directly from his/her actions. This process would work just the same even if there were no groups at all.

This process does indeed account for much cooperation, but this is not "altruism", or at least not altruism as most people know it. In these models, a purely self-interested act (e.g. acquiring food) is redefined as being "altruistic" simply because others happen to benefit more. This point is nicely illustrated in Reeve's (2000) clear exposure of the mathematics underlying group selection and the redefinition of "altruism". This redefinition of "altruism" deviates considerably from the common usage of the term and requires considerable chutzpah to say in a straight face – a cynical author might even call it "sleight of mouth". In these cases, acts that are individually-beneficial but benefit others *more* could still be selected for as long as there is some degree of "global competition" (i.e. relative to the global population) rather than competition being completely local (i.e. relative to the local group, see West et al. 2006). In fact, any benefits to the group could be completely incidental because they are not necessary for these models – all that matters is

that the actors themselves benefit from the provision of public goods (Harpending, 1998). As such, this type of selective pressure is unlikely to select for group-level adaptations (i.e. cognitive mechanisms or psychological propensities that exist solely because they are "good for the group") – a point recently proven mathematically by Gardner and Grafen (2009).

Because of this redefinition of "altruism" and the reliance on helpers benefiting themselves (relative to the broader population), this first type of "group selection" is essentially a misnamed form of individual-level selection. The problem is not with the models or the ideas, but with the highly (and possibly deliberately) misleading name. It would be better to simply call this inclusive fitness maximization, while recognizing that the scale of competition (global versus local competition) is important.

### 5.1.2. Decreasing Relative and Absolute Fitness – But How Do Groups Form?

In the second type of situation, individuals do something that decreases their fitness relative to the *global* population in order to confer benefits on others, which fits more common definitions of the word "altruism". Groups with many such cooperators will indeed benefit more than groups with many non-cooperators. However, we must ask how such non-random groups form and how cooperators manage to assort themselves such that they interact more with other cooperators than non-cooperators do. Without such assortment, cooperators will be outcompeted by defectors unless there are direct benefits for cooperating as in section 5.1.1 (Fletcher and Doebeli, 2009).

The most common cause of such assortment is if most group members evolved from a common ancestor, and low gene flow has kept them together and prevented the groups of cooperators from being overrun by non-cooperators. In such cases with low gene flow, it is *kinship* that drives the evolution of altruism: altruists help group members, who because of common descent just happen to be more closely related to the altruist than to the average group member is (and thus are more likely to carry copies of any genes that influence altruism). Thus, this type of "group selection" is nothing more than another way of describing inclusive fitness maximization, which has long been recognized as a cornerstone of social evolution (see section 2). This point has long recognized by those willing to slog through the mathematics underlying so-called group selection (e.g. Foster et al., 2006; Reeve, 2000; Sober and Wilson, 1998; West et al., 2007), but has sometimes

been apparently forgotten or conveniently neglected by some authors when promoting group selection.

Of course, common descent is not the only way that altruists can assort with one another (Fletcher and Doebeli, 2009). For example, individuals could choose their social partners, such that altruists choose to associate with other altruists, forming groups of altruists. When this happens, "altruism" becomes directly beneficial to the "altruist" because it increases his/her later likelihood of associating with others who confer benefits upon group members. Thus, the behavior isn't really altruistic anymore – cooperative yes, generous sure, but "altruistic" no, because there are direct benefits for being generous. Similarly, cooperators may use reciprocity such that generous people elicit generous reactions from others and selfish people elicit selfish reactions from others. Thus, by being generous, a person maintains the cooperation in his/her group, and thus he/she benefits from being generous. In both of these situations, the generous behavior is directly beneficial to oneself, so it can evolve by individual selection without regards to what happens to the rest of the group (see section 5.1.1). As such, the term "group selection" is a misnomer for these types of assortment because it relies on direct benefits to self rather than on the effects upon other group members. Section 6.2 will cover assortment in more detail.

### 5.1.3. "Cultural Group Selection"

In a third type of model, cooperation spreads as less successful groups adopt the cooperative and punitive norms of more successful groups (Boyd and Richerson, 2002; Guzman et al., 2007). Any norms that are good for groups that hold them – such as within-group cooperation and punishment of non-cooperators – are likely to do better than norms that are detrimental for groups. In these models of "cultural group selection", it is cultural norms that are being spread at the expense of other cultural norms, rather than genetically-influenced strategies spreading at the expense of other genetically influenced strategies (Richerson and Boyd, 2005). Cooperative norms could spread either by complete replacement of groups with unsuccessful norms (i.e. non-cooperation) by more cooperative and successful groups, partial replacement of less successful groups whereupon the remaining members adopt the norms of the successful invading group, or pure imitation where less successful groups are not replaced at all but instead willingly adopt the more functional norms of more successful groups (i.e. cooperation or self-sacrifice).

These models are less affected by people moving between groups because people tend to conform to the norms of their groups (Henrich and Boyd, 2001; Henrich and Gil-White, 2001), thus maintaining the relative homogeneity of behaviour within groups. As such, these models are rather more plausible than models of group selection that rely on genetic selection alone. One recent study showed that groups who could punish free-riders earned more than groups who could not, and people tended to leave the latter to join the former, such that the presence of punishment tended to increase (Gürerk et al., 2006).

Despite the potential usefulness of such models, once again it is a misnomer to add the word "group" to "cultural group selection". When discussing the spread of cultural variants, the existence of a "group" is unnecessary, as is competition between rival cultural groups. A behavioral norm can spread if it does better than alternative norms, regardless of whether it affects groups or individuals. Natural selection will sometimes act on group-level norms such as institutions, but this is not necessary for selection of different cultural variants (for an example that was very recent at the time of writing, see Strimling et al, 2009). Since group-level effects are not necessary, these models are better labeled "cultural selection" (i.e. natural selection acting on culture itself) as opposed to "cultural group selection" (which implies selection acting on *groups* who have culture).

In addition to having a misleading name, there are potential problems with using cultural group selection models to explain the evolution of individually-costly behaviors such as generosity and punishment. It is currently unclear how cooperation, punishment, and rewarding become common in the first place within groups in such group selection models. Conformist-transmission (Henrich and Boyd, 2001) and cultural group selection (Boyd and 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 (or a similar cultural process) would have to be very strong to overcome the selection against generosity and "altruistic punishment" and make them the most common behaviours. Prestige-based imitation (imitating the most successful group members, Henrich and Gil-White, 2001) alone cannot account for the presence of cooperation and punishment unless

generous and punitive individuals already have high status (Henrich and Boyd, 2001), and this begs the question of why they would tend to have high status. 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 myriad 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 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.

## 5.2. DOES ANONYMOUS GENEROSITY SUPPORT GROUP SELECTION?

Some researchers argue that people are generous and punitive even in anonymous one-round laboratory tasks where there are no incentives to appear, so this therefore provides evidence that the sentiments causing such behaviour could not have evolved due to the reputational benefits for such behaviour and instead must have evolved due to group-level selection pressures (Fehr et al., 2002; Fehr and Henrich, 2003; Gintis, Bowles, Boyd, and Fehr, 2003). This viewpoint has been criticized on the grounds that generosity towards unrelated strangers in anonymous laboratory tasks is also maladaptive from a group selectionist perspective because the relevant "group" is far too big to be selected by group selection (Barclay, 2005; Burnham and Johnson, 2005), given that people will be generous towards people in other parts of the country or even other countries (e.g. DeBruine, 2002; Eckel and Wilson, 2004; Krupp et al., 2008; Yamagishi et al., 2005), such that group selection provides no better an account of this maladaptive

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.

behaviour than do reputational accounts. Also, this particular group selection viewpoint assumes that people are making perfectly rational decisions in the experimental tasks (Hagen and Hammerstein, 2006), which is a very large assumption that is not necessarily correct.

Perhaps the most powerful argument against this group-selectionist account of generosity in anonymous tasks is that it is simply 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 and Fehr, 2003). If there have been past selective pressures favouring generosity, 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 a person receives some sort of emotional reward or positive feeling from cooperating or punishing non-cooperation (de Quervain et al., 2004), then they are likely to receive that reward or have that feeling whether they are in a laboratory or not. Thus, a person may very well be generous or punitive because he/she enjoys being generous or punitive, even if he/she consciously believes that his/her actions are unknown to others. However, the presence of such a psychological mechanism says little about the selection pressures that would have caused such a mechanism to exist in the first place – if a genuine concern for others evolves because people with that concern tend to receive reputational benefits, then it will cause generous behaviour in situations both with and without reputational incentives, but only generosity in the former will actually bring the benefits that select for the behaviour. As long as the former are more frequent and/or carry greater fitness costs/benefits, then the mechanisms can be adaptive *on average*. Having such genuine concern may be beneficial because it commits oneself to being nice, which will change others behaviour towards oneself (Frank, 1988).

Even if any psychological mechanisms that evolved for reputation management do facultatively adjust generosity according to the presence of reputational incentives, one might expect them to err on the side of assuming non-anonymity, given that one's perceptions of anonymity can be wrong (Frank, 1988; Johnson and Bering, 2006). Consistent with this, there is evidence that even in anonymous situations, participants display higher levels of trust and generosity when they are presented with subtle cues of kinship

with fellow participants (DeBruine, 2002; Krupp et al., 2008) or photographs or sketches of watchful eyes (Bateson et al., 2006; Burnham and Hare, 2007; Haley and Fessler, 2005), even though participants are fully aware that their decisions are anonymous and that they are not interacting with kin.

*Chapter 6*

# INDIVIDUAL-LEVEL BENEFITS FOR PROVIDING PUBLIC GOODS AND INCENTIVES

If generous people (including incentive-providers) receive individual benefits for their acts, then this could make up for the cost of such behaviours and select for generous and punitive sentiments. Once generosity, 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 generosity between groups, the possibility of group selection does not necessitate that generosity is a group-level adaptation, as Wilson has claimed (1998; Sober and Wilson, 1998) – see Gardner and Grafen (2009) for a mathematical proof of this.

## 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, generosity 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, and those who do not provide public goods or reward cooperators are simply excluded from this system of generalized exchange.

Panchanathan and Boyd (2004) demonstrated this mathematically with a 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. It is not a new idea that norms can support behavior, but the trick is how a particular action comes to be seen as "good" and thus deserving of reward in indirect reciprocity. If an act obviously benefits others, then it makes sense that it will be seen as good. For some public goods where the benefits to others are not immediately obvious (e.g. the fight against climate change or other such long-term goods), it may be harder to link the public good provision with indirect reciprocity because some others may not realize the importance of the act. Nevertheless, once providing the public good is seen as being worthy of reward – for example if then public is informed about the importance of the good (Milinski et al., 2006) – then it can sustained by indirect reciprocity.

Behavioral experiments support the idea that indirect reciprocity can support the provision of public goods. Milinski and colleagues (2002a, b; Semmann et al., 2004) found that people rewarded those who cooperated in public goods games, even those in completely different groups (Semmann et al., 2005) In addition to discriminating against those who do not provide public goods, people also discriminate against those who do not reward public good providers: Kiyonari and Barclay (2008) 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. More recently, Rockenbach and Milinski (2006) showed that the presence of indirect reciprocity reduces the need for punishment (and therefore the cost to punishers) because there exists another, cheaper way to maintain the provision of public goods (Rockenbach and Milinski, 2006). Milinski and colleagues have shown that these systems of indirect reciprocity can be effective tools in the fight against climate change because they cause people to cooperate to prevent climate change (Milinski et al., 2006, 2008). This is most effective when people are well-informed about the dangers of climate change, i.e. when it becomes more obvious that cooperation is important and benefits others.

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 and Brewer, 1983), are more punitive of norm violations against ingroup members than outgroup members (Bernhard et al., 2006), and they accord ingroup members more money in monetary-sharing experiments (e.g. Billig and Tajfel, 1973; Tajfel, Billig, Bundy, and 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 an adaptation that exists because it is good for the group. 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 and Kiyonari, 2000). Ingroup favouritism disappears when a person's payoff cannot be affected by others' decisions (Karp, Jin, Yamagishi, and 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 and Kiyonari, 2000). Thus, it appears that expectations of reciprocity account for ingroup favouritism better than a hypothesized group-level adaptation would.

## 6.2. ASSORTATIVE INTERACTIONS, PARTNER CHOICE, AND COMPETITIVE GENEROSITY

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. Others have found similar results using other types of social dilemmas (Burlando and Guala, 2005; Gunnthorsdottir et al., 2000; Yang et al., 2007), and the threat of expulsion from such cooperative groups results in higher cooperation within the groups (Cinyabuguma et al., 2005). As for endogenously determined relationships, Sheldon, Skaggs Sheldon, and 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.

When people can choose with whom to interact, there is competition for access to the best social partners including the most cooperative partners, such that individuals may use generosity to compete over access to potential partners (Roberts, 1998; Seyfarth, 1977). Such generosity can increase the number or quality of partners available to a person because others will benefit from associating with people who provide such benefits. As some individuals start using generosity to attract partners, it gives others an incentive to be *even more* generous so that *they* can be the one to attract the best partners. This process has been called "competitive altruism"[1] or "competitive generosity", and occurs when individuals try to act not merely generously, but *more* generously than others. Competitive generosity may account for lavish public generosity like the provision of feasts including New Guinean mokas and Kwakiutl potlatches (Boone, 1998; Goldman, 1937; Strathern, 1979; Smith and Bliege Bird, 2000). Such competitive generosity can occur as long as the generosity carries some useful information about a person, such as their resources, abilities, or cooperative intent – this point will be covered in the next section on costly signaling. This competition need not be conscious, and could occur over evolutionary time as competition over partners causes an escalation of generosity in the population.

1 Remember that some fields use the term "altruism" to refer simply to costly behaviours that benefit others, without reference to the specific psychological states or motivations involved.

Many experiments show that people are more generous when others are watching (e.g. Barclay, 2004; Hardy and Van Vugt, 2006; Rege and Telle, 2004). This by itself is not competitive generosity unless there is actual competition over partners. In studies that institute competition over partners in at least one experimental condition, people are most generous when their potential partners get to choose with whom to interact such that people must use competitive generosity to compete for those partnerships (Barclay, 2004; Barclay and Willer, 2007). This demonstrates that people will strive to not just be generous, but *more* generous than their competitors.

This competitive generosity is part of a market for social partners (Seyfarth, 1977, see also Noë and Hammerstein, 1994, 1995) where the best cooperators will end up with each other and the rest have to make do with whoever is left. This process is analogous to a marriage market, except that it applies to other types of relationships like friendships and coalition partners as well (for a review, see Barclay, in preparation). This process of partner choice based on observed levels of generosity can result in a "runaway" escalation of generosity towards very high levels (McNamara et al., 2008; Nesse, 2007).

Although groups of cooperators do better than groups of non-cooperators when there are assortative interactions, this need not be considered group selection as some authors (e.g. 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 cooperate in order 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.

## 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 generosity) 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.

### 6.3.1. Generosity as a Costly Signal of Abilities and Resources

Costly signaling theory can be applied to generosity. Generosity, by definition, is costly to the generous individual. However, the same generous 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 and Rohner, 1970; but see Drucker and Heizer, 1967). For example, when billionaires such as 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

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

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 have a higher risk of drowning if they tried, so the act is less costly to me than it would be to a weaker swimmer. In both these cases, generous individuals can benefit from having others know about their underlying quality (wealth or physical ability in these examples), and observers benefit from knowing the giver's 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 (Doutrelant and Covas, 2007; Lotem, Wagner, and Balshine-Earn, 1999). For example, I may jump into the river because it is my baby and my genetic interest in the child's welfare causes me to be concerned enough to jump in, 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 helping less related or even unrelated babies, or for an increased level of generosity towards those babies, in order to demonstrate my abilities. Similarly, signaling benefits can create a selective pressure for generosity not only towards reciprocators, but also towards people who are unlikely or unable to reciprocate (Lotem, Fishmann, and 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 generous 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, and Bird, 2001; Hawkes and Bliege Bird, 2002; Smith, Bliege Bird, and 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 and Smith, 2005).

Although many different kinds of costly behaviour could be used to signal quality (for a review, see Bliege Bird and 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 provided (Hawkes and Bliege Bird, 2002; Gintis et al., 2001; Smith and Bliege Bird, 2000). Thus, prosocial signals can attract a larger audience per unit of effort than other costly signals (Boone, 1998; Smith and Bliege Bird, 2005). If individuals are competing with each other to attract the best mates and allies and to deter others, and they are using generosity as a costly signal of quality, then they may compete to be the most generous 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 generosity than do others who are not as generous. Some primate researchers have suggested that baboons compete to groom the highest-ranking group members (Barrett et al., 2000), and humans use generosity to compete for access to social partnerships (Barclay, 2004; Barclay and Willer, 2007).

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 and 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 and 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 and 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 such as 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 successful 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.

### 6.3.2. Generosity as aCostly Signal of Cooperative Intent

Clearly, not all generous acts are sufficiently difficult or costly such that they could be costly signals of abilities or resources. Some generous 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 generosity is not worth the cost for those who intend to defect on cooperative partners (Smith, 2003). Smith and Bliege Bird (2005) 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 and Houston, 2002; McNamara et al., 2009). 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 directly tested whether public generosity signals a willingness to cooperate with others. 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.

Albert, Güth, Kirchler, and Maciejovsky (2007) 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 (2007) 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 generous people were not more trustworthy overall, but were more trustworthy towards other generous people, so the giving could be an honest signal of cooperative intent towards other generous people. These results do not show that people actively signaled their generosity, but did show that people responded to it as if it were a signal.

Much research is consistent with the idea that generosity functions like a costly signal of cooperative intent. For example, people cooperate more with and entrust more money to those who have a history of generous behavior (Barclay, 2004, 2006; Barclay and Willer, 2007; Wedekind and Braithwaite, 2002). Heroic risk-takers (and other people who help others) are more attractive for long-term relationships than are non-helpers and non-heroes (Barclay, 2010; Farthing, 2005), and may even be attractive to women for slightly shorter relationships (Barclay, 2010). People respond to the presence of these benefits by being more cooperative when others are aware of their decisions (Barclay, 2004; Barclay and Willer, 2007; Hardy and Van Vugt, 2006; Hoffman et al., 1994; Rege and Telle, 2004); the mere presence of eyes or stylized eyespots is enough to induce higher levels of generosity in laboratory (Burnham and Hare, 2007; Haley and Fessler, 2005) and naturalistic settings (Bateson et al., 2006).

People often do not directly observe others' actions but can instead use information gathered from others. In an experiment by Keser (2003), people played a series of trust games (from Berg, Dickhaut, and 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 entrust 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 (2003) 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. Other research shows that people will use verbal information from others (i.e. gossip) even when they can use their own first-hand observation (Sommerfeld et al., 2007). This opens up the possibility that gossip about others' cooperativeness can be used manipulatively, and we would expect people to have cognitive strategies for dealing with gossip and assessing the validity of social information (e.g. Hess and Hagen, 2006).

Gossip and social information are also important in other accounts of the evolution of generosity, such as indirect reciprocity, and some of these aforementioned effects are also consistent with indirect reciprocity. Indeed, it can be difficult to differentiate between indirect reciprocity accounts and costly signaling of cooperative intent because they are both reputation models and both rely on reputational benefits accruing to generous individuals. It may be possible that there is no inherent difference between these types of models, and that cooperation and rewarding within a system of indirect reciprocity could function as costly signals of future willingness to contribute to that system, such that indirect reciprocity only works because of costly signaling. Nevertheless, there may be ways to separate these accounts. A crucial point is that under costly signaling, observers glean valuable information from observing others' cooperation, and then act out of their own interest in a way that is beneficial to the signaler (e.g. allying with someone with many resources or trusting someone with cooperative intent). Thus, when people act against their vested interest to help generous individuals, it suggests indirect reciprocity, whereas when people act on information according to their vested interests in dealing with generous individuals, it is more supportive of costly signaling. For example, if there are cases where generous individuals are trusted (which is in the truster's vested interest) but not rewarded (which is not in the rewarder's vested interest if unobserved), or where people continue to act favourably towards generous individuals long after those individuals had been compensated for their efforts, or where people behave in a way that is beneficial to themselves but neutral towards generous individuals, then this would support costly signaling accounts more than indirect reciprocity. Consistent with these predictions, Barclay (2004, 2006) found that people tend to trust public good providers and justified punishers, but did not seem to reciprocate more trust to them, although situational factors and lower statistical power could have overwhelmed the latter effect (see Barclay, 2004, for a discussion of trust versus reward).

### 6.3.3. Punishment of Free-Riders 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 and 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 and Kinsey, 2002; Wiessner, 2003). It would be interesting to see whether gaining or losing status affects people willingness to provide other forms of punishment.

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 (Chagnon, 1997; Daly and Wilson, 1988; Johnstone and Bshary, 2004). This is allegedly occurring in "cultures of honour" in places such as the southern United States (Cohen, Nisbett, Bowdle, and 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 (Daly and Wilson, 1988).

Similarly, when people sanction free-riders in a group, they may be signaling an unwillingness to be cheated. The psychological causes of punishment need not necessarily be to deter future transgressions (e.g. Carlsmith, 2002), but the emotions that cause punishment may exist because punishment deters transgressions against oneself. Observers would 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, and Sigmund, 2003; Hauert, Haiden, and Sigmund, 2004; Sigmund, Hauert, and 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, and Sigmund, 2000). Experimental evidence suggests that people are more likely to reject unfair offers when they can acquire a reputation for doing so (Fehr and 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., 2007).

In support of these hypotheses, Barclay (submitted) has shown that when people play "Taking Games" in which they can take money from others and pay to retaliate against such taking, participants who provide punishment of free-riders in cooperative games tend to have less money taken from them than do non-punisher. This suggests that people are less willing to exploit punishers because of the potential costs of exploiting those who have demonstrated a dislike of unfairness. Just as men are generally more concerned with deterrence and appearances of toughness than are women (Daly and Wilson, 1988), men tend to provide more punishment of free-riders than do women (Barclay, 2006).

Finally, punishment of non-cooperators 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 dislike 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 generosity 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).

To test this prediction, Barclay (2006) had participants play public goods games with punishment (e.g. Fehr and Gächter, 2002) followed by "Trust Games" (Berg et al., 1995), and found that participants entrusted more money

to punishers than to non-punishers, provided that the punishment was "justified" punishment of low contributors rather than unjustified punishment of average or higher-than-average cooperators. Barclay also shows that this trust of punishers only occurs if participants have sufficient exposure to free-riders and punishment, such that the justification for punishment is clear. As predicted by a costly signaling account of these reputational benefits, Nelissen (2008) demonstrated that people's trust of punishers does indeed depend upon the cost of punishment: people only trusted more punishers than non-punishers and chose the former as cooperative partners when the punisher incurred sufficient cost upon him/herself to punish.

If punishment is a signal of quality, status, cooperative intent, or unwillingness to be cheated, then others should 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 (Barclay, 2006, submitted). If such reputational benefits translated into tangible benefits in ancestral environments, then this could explain the existence of punitive sentiments.

*Chapter 7*

# MULTIPLE REPUTATIONAL EFFECTS OF GENEROSITY

I have outlined some different types of reputational benefits that individuals may receive after they perform generous behaviour and that might cause the natural selection and/or learning of generous sentiment, but none of the benefits for generosity are necessarily mutually exclusive. For example, people "showing off" by helping others could receive more help when in need (indirect reciprocity), be chosen more often as cooperative partners (assortative interactions) or 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). Other group members would give aid to these show-offs in order to keep in good standing with others (indirect reciprocity) and to remain close enough to them to receive the benefits of those persons' generosity (assortative interactions), mate with them in order to have high quality mates or defer to them in order to avoid costly competition with a superior competitor (costly signaling of individual quality), and trust them because they are deemed less likely to cheat a cooperative partner (costly signaling of cooperative intent). Once a group has a system whereby individuals receive reputational benefits for the generosity, it will likely result in higher levels of cooperation, such that that group will do better than other groups and spread the genes and/or cultural norms that cause such generosity by simple proliferation of genes/norms rather than by reputational benefits (group selection and cultural group selection).

All of that being said, anything that increases the benefits to a generous individual 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 generosity increase, we would expect signals of generosity to become more costly in order to maintain their honesty and audiences to become more skeptical of the intent of such signals (Barclay and Willer, 2007). Berman and Laitin (2008) 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 generosity clearly predicts an escalation in the degree of generosity displayed when the potential benefits of generous signaling increase. Costly signaling and assortative interactions both predict an escalation of generosity in response to increased generosity by others, in that people 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, which over evolutionary time can lead to an escalation of cooperation and morality (Miller, 2007; Nesse, 2007). Thus, the different theories do make different predictions about generosity. While the mechanisms that select for generous sentiment may all be occurring at the same time, some may be more important in some situations than others.

*Chapter 8*

# CONCLUSIONS

Humans help non-kin more than perhaps any other species on earth. Evolutionary accounts attempt to explain why humans possess psychological mechanisms that cause this generosity. Whether these psychological mechanisms are emotions, cognitive algorithms, or whatnot, what is the function of possessing them? Why did those psychological mechanisms evolve and/or why was generous behavior learned despite the costs of helping others – what selective pressures and social forces compensate generous individuals for the costs of their acts? Reputational benefits provide a strong selective pressure for generosity because generous people may receive reciprocal aid when they are in need, higher rates of rewards and lower punishment, preferential access to groups and social partners, or they may benefit from signaling important traits about themselves. Of course, this does not mean that people *intend* to receive such benefits when they act generously. Instead, people who *do* act generously will tend to receive these benefits, which reinforces the behaviors, and if this occurs over evolutionary time will tend to cause natural selection for whatever sentiments produce the behavior.

This book reviews some of the theories and empirical studies showing that reputational benefits can support generosity. However, this review does not intend to imply that reputational benefits are the *only* type of benefits for generous behavior – sometimes people benefit directly from performing group-beneficial acts because they benefit from the act being performed (e.g. Doebeli and Hauert, 2005; West et al., 2006) or because they have a vested interest in others' wellbeing (Roberts, 2005; Tooby and Cosmides, 1996). Nevertheless, social pressures are extraordinarily important for humans, such

that some type of reputational benefits could very well be the most important factor supporting human generosity. Furthermore, it may be easier to manipulate the presence of reputational benefits than it would be to directly manipulate other factors, whether they be factors related to functional causes of generosity (e.g. the degree of mixing between groups as it affects group selection) or the psychological causes of altruism (e.g. feelings of oneness with others). As such, reputational benefits and are well worth investigating as potential means to increase the amount of generosity and cooperation in the world.

# REFERENCES


Albert, M., Güth, W., Kirchler, E., and Maciejovsky, B. (2007). Are we nice(r) to nice(r) people? An experimental analysis. *Experimental Economics*, *10*, 53-69.

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

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. (1986). An evolutionary approach to norms. *American Political Science Review*, 80, 1095-1111.

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

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

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

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

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

Barclay, P. (2005). Reputational benefits of altruism and altruistic punishment. Unpublished doctoral dissertation, McMaster University, Canada.

Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior*, 27, 325-344.

Barclay, P. (2008). Enhanced recognition of defectors depends on their rarity. Cognition, 107, 817-828.

Barclay, P. (2010). Altruism as a courtship display: Some effects of third-party generosity on audience perceptions. *British Journal of Psychology*, *101*, 123-135.

Barclay, P. (submitted). "Don't mess with the enforcer": Deterrence as an individual-level benefit for punishing free-riders.

Barclay, P. (in preparation). Biological markets, partner choice, and the evolution of competitive altruism.

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

Barclay, P., and Lalumière, M. L. (2006). Do people differentially remember cheaters? Human Nature, 17(1), 98-113.

Barclay, P., and Willer, R. (2007). Partner choices creates competitive altruism in humans. Proceedings of the Royal Society of London B, 274, 749-753.

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., and 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., and Hill, R. A. (2000). Female baboons do not raise the stakes but they give as good as they get. Animal Behaviour, 59, 763-770.

Bateson, M., Nettle, D., and Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. Biology Letters, 2, 412-414.

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

Berman, E., and Laitin, D. D. (2008). Religion, terrorism and public goods: testing the club model. *Journal of Public E*conomics, 92, 1942-1967.

Bernhard, H., Fischbacher, U., and Fehr, E. (2006). Parochial altruism in humans. Nature, 442, 912-915.

Betzig, L. (1988). Redistribution: equity or exploitation? In Human Reproductive Behaviour: A Darwinian Perspective (Eds. L Betzig, M.

Borgerhoff Mulder, and P. Turke), pp. 49-63. Cambridge, UK: Cambridge University Press.

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

Billig, M., and 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., and Kushnik, G. C. (2002). Risk and reciprocity in Meriam food sharing. Evolution and Human Behavior, 23, 297-321.

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

Bliege Bird, R., Smith, E. A., and 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., and Ockenfels, A. (2005). Cooperation among strangers with limited information about reputation. *Journal of Public Economics*, 89, 1457-1468.

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

Bowles, S. (2006). Group competition, reproductive leveling, and the evolution of human altruism. Science, 314, 1569-1572.

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

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

Boyd, R., and 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., and 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.

Bshary, R., and Bergmüller, R. (2008). Distinguishing four fundamental approaches to the evolution of helping. *Journal of Evolutionary Biology*, 21, 405-420.

Burlando, R. M., and Guala, F. (2005). Heterogeneous agents in public goods experiments. Experimental Economics, 8, 35-54.

Burnham, T., and Hare, B. (2007). Engineering human cooperation: does involuntary neural activation increase public goods contributions? Human Nature, 18, 88-108.

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.

Chagnon, N. (1997. Yanomamö (5th ed.). Wadsworth.

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

Choi, J.-K., and Bowles, S. (2007). The coevolution of parochial altruism and war. Science, 318, 636-640.

Cinyabuguma, M., Page, T., and Putterman, L. (2005). Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics*, 89, 1421-1435.

Cinyabuguma, M., Page, T. and Putterman, L. (2006). Can second-order punishment deter perverse punishment? Experimental Economics, 9, 265-279.

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

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

Cohen, D., Nisbett, R. E., Bowdle, B. F., and 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., and 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., and 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., and Tooby, J. (1992). Cognitive Adaptations for Social Exchange. In The Adapted Mind: Evolutionary Psychology and the Generation of Culture (eds. J. Barkow, L. Cosmides, and 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., and Wilson, M. (1983). Sex, Evolution, and Behavior. Belmont, CA: Wadsworth Publishing Company.

Daly, M., and 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., and 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., and 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.

Denant-Boemont, L., Masclet, D., and Noussair, C. N. (2006). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. Economic Theory, 33, 145-167.

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

Doebeli, M., and Hauert, C. (2005). Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game. Ecology Letters, 8, 748-766.

Doutrelant, C., and Covas, R. (2007). Helping has signaling characteristics in a cooperatively breeding bird. Animal Behaviour, 74, 739-747.

Drucker, P., and 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 and Company.

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

Dwyer, P. D., and 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., and 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.

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

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

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

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

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

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

Fehr, E., and 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., and 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., and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. Economics Letters, 71, 397-404.

Fletcher, J. A., and Doebeli, M. (2009). A simple and general explanation for the evolution of altruism. Proceedings of the Royal Society of London B, 276, 13-19.

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

Foster, K. R., Wenseleers, T., Ratnieks, F. L. W., and Queller, D. C. (2006). There is nothing wrong with inclusive fitness. Trends in Ecology and Evolution, 21(11), 599-600.

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

Gächter, S., Renner, E., and Sefton, M. (2008). The long-run benefits of punishment. Science, 322, 1510.

Gardner, A., and Grafen, A. (2009). Capturing the superorganism: a formal theory of group adaptation. *Journal of Evolutionary Biology*, 22, 659-671.

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

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

Gintis, H., Smith, E. A., and 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., and Ameden, H. (2000). Excluding free-riders improves reciprocity and promotes the private provision of public goods. Working Paper.

Gürerk, Ö., Irlenbusch, B., and Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. Science, 312, 108-111.

Gurven, M., Allen-Arave, W., Hill, K., and 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., and 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., and Lyles, R. (2000). Food transfers among Hiwi foragers of Venezuala: tests of reciprocity. Human Ecology, 28, 171-218.

Guzman, R. A., Rodríguez-Sickert, C., and Rowthorn, R. (2007). When in Rome, do as the Romans do: the coevolution of altruistic punishment, conformist learning, and cooperation. Evolution and Human Behavior, 28, 112-117.

Haley, K. J., and 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 and 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.

Hardy, C., and Van Vugt, M. (2006). Giving for glory in social dilemmas: The competitive altruism hypothesis. Personality and Social Psychology Bulletin, 32, 1402-1413.

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., and 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., and 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.

Hauser, M. D., and Fehr, E. (2007). An incentive solution to the peer review problem. PLoS Biology, 5(4), 0703.

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., and 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., and Blurton Jones, N. G. (2001a). Hadza meat sharing. Evolution and Human Behavior, 22, 113-142.

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

Henrich, J., and 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., and 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., and 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., and 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., and Tracer, D. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. Behavioral and Brain Sciences, 28, 795-855.

Henrich, J., and 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.

Herrmann, B., Thöni, C., and Gächter, S. (2008). Antisocial punishment across societies. Science, 319, 1362-1367.

Hess, N. H., Hagen, E. H. (2006). Psychological adaptations for assessing gossip veracity. Human Nature, 17(3), 337-354.

Hill, K., and 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, and P. Turke), pp. 277-289. Cambridge, UK: Cambridge University Press.

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

Johnson, D., and Bering, J. (2006). Hand of God, mind of man: punishment and cognition in the evolution of cooperation. Evolutionary Psychology, 4, 219-233.

Johnstone, R. A., and Bshary, R. (2004). Evolution of spite through indirect reciprocity. Proceedings: Biological Sciences, 271, 1917-1922.

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

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

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

Kiyonari, T., and Barclay, P. (2008). Cooperation in social dilemmas: free-riding may be thwarted by second-order reward rather than by punishment. *Journal of Personality and Social Psychology*, 95(4), 826-842.

Kiyonari, T., Shimoma, E., and 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., and Parks, C. D. (1995). Interpersonal relations: mixed-motive interaction. Annual Review of Psychology, 46, 183-207.

Krams, I., Krama, T., Igaune, K., Mänd, R. (2008). Experimental evidence of reciprocal altruism in the pied flycatcher. Behavioral Ecology and Sociobiology, 62, 599-605.

Krupp, D. B., DeBruine, L. M., and Barclay, P. (2008). A cue of kinship promotes cooperation for the public good. Evolution and Human Behavior, 29, 49-55.

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

Kurzban, R., DeScioli, P., and O'Brien, E. (2007). Audience effects on moralistic punishment. Evolution and Human Behavior, 28, 75-84.

Lahti, D.C., and Weinstein, B.S. (2005). The better angels of our nature: group stability and the evolution of moral tension. Evolution and Human Behavior, 26(1), 47-63.

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

Lehmann, L., Foster, K.R., Borenstein, E., and Feldman, M. W. (2008). Social and individual learning of helping in humans and other species. Trends in Ecology and Evolution, 23(12), 664-671.

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

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

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

Majolo, B., Ames, K., Brumpton, R., Garratt, R., Hall, K., and Wilson, N. (2006). Human friendship favours cooperation in the Iterated Prisoner's Dilemma. Behaviour, 143, 1383-1395.

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

Mashima, R., and 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., and Price, G. R. (1973). The logic of animal conflict. Science, 246, 15-18.

McCabe, K. A., Rigdon, M. L., and 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., and 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., Frohmage, L., and Houston, A. I. (2008). The coevolution of choosiness and cooperation. Nature, 451, 189-192.

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

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

McNamara, J. M., Stephens, P. A., Dall, S. R. X., and Houston, A. I. (2009). Evolution of trust and trustworthiness: social awareness favours

personality differences. Proceedings of the Royal Society of London B, 276, 605-613.

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., and Krage, M. (1996). Enhanced memory for faces of cheaters. Evolution and Human Behavior, 17, 119-128.

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

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

Milinski, M., Pfluger, D., Külling, D., and 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., and Krambeck, H.-J. (2001). Cooperation through indirect reciprocity: image scoring or standing strategy? Proceedings: Biological Sciences, 268, 2495-2501.

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

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

Milinski, M., Semmann, D., Krambeck, H.-J., and Marotzke, J. (2006). Stabilizing the Earth's climate is not a losing game: supporting evidence from public goods experiments. PNAS, 103(11), 394-3998.

Milinski, M., Sommerfeld, R. D., Krambeck, H.-J., Reed, F. A., and Marotzke, J. (2008). The collective-risk social dilemma and the prevention of simulated dangerous climate change. PNAS, 105(7), 2291-2294.

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

Miller, G. F. (2007). Sexual selection for moral virtues. The Quarterly Review of Biology, 82(2), 97-125.

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.

Nelissen, R. M. A. (2008). The price you pay: cost-dependent reputation effects of altruistic punishment. Evolution and Human Behavior, 29, 242-248.

Nesse, R. M. (2007). Runaway social selection for displays of partner value and altruism. Biological Theory, 2(2), 143-155.

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.

Noë, R., and Hammerstein, P. (1994). Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. Behavioral Ecology and Sociobiology, 35, 1-11.

Noë, R., and Hammerstein, P. (1995). Biological markets. Trends in Ecology and Evolution, 10, 336-339.

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

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

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

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

Nowak, M. A., and Sigmund, K. (2005). Evolution of indirect reciprocity. Nature, 437, 1291-1298.

Ohtsuki, H., and Iwasa, Y. (2004). How should we define goodness? Reputation dynamics in indirect reciprocity. *Journal of Theoretical Biology*, 231, 107-120.

Ohtsuki, H., and Iwasa, Y. (2006). The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology*, 239, 435-444.

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

Olendorf, R., Getty, T., and 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., and 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., and 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., and Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. Nature, 432, 499-502.

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

Petrinovich, L., O'Neill, P., and 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.

Puurtinen, M., and Mappes, T. (2009). Between-group competition and human cooperation. Proceedings of the Royal Society of London, 276, 355-360.

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

Reeve, H. K. (2000). Multi-level selection and human cooperation. Evolution and Human Behavior, 21, 65-72.

Richerson, P. J., and Boyd, R. (2005). Not By Genes Alone: How Culture Transformed Human Evolution. Chicago, IL: University of Chicago Press.

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

Roberts, G. (2005). Cooperation through interdependence. Animal Behaviour, 70, 901-908.

Roberts, G. (2008). Evolution of direct and indirect reciprocity. Proceedings: Biological Sciences, 275, 173-179.

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

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

Rockenbach, B., and Milinski, M. (2006). The efficient interaction of indirect reciprocity and costly punishment. Nature, 444, 718-723.

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

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

Sefton, M., Shupp, R., and 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., and Schram, A. (2006). Social status and group norms: indirect reciprocity in a helping experiment. European Economic Review, 50, 581-602.

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

Semmann, D., Krambeck, H.-J., and Milinski, M. (2005). Behavioral Ecology and Sociobiology, 57, 611-616.

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., and Osbaldiston, R. (2000). Prosocial values and group assortation. Human Nature, 11, 387-404.

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

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

Smeesters, D., Warlop, L., Van Avermaet, E., Corneille, O., and 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., and Bliege Bird, R. (2000). Turtle hunting and tombstone opening: public generosity as costly signaling. Evolution and Human Behavior, 21, 245-262.

Smith, E. A., and Bliege Bird, R. (2005). Costly signaling and cooperative behavior. In Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life (Eds. H. Gintis, S. Bowles, R. Boyd, and E. Fehr), pp. 115-148. Cambridge, MA: MIT Press.

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

Sober, E., and 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., and Alcorta, C. (2003). Signaling, solidarity, and the sacred: the evolution of religious behavior. Evolutionary Anthropology, 12, 264-274.

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

Stevens, J., and Hauser, M. (2004) Why be nice? Psychological constraints on the evolution of cooperation. Trends in Cognitive Science, 8(2), 60-65.

Stewart-Williams, S. (2007). Altruism among kin vs. nonkin: effects of cost of help and reciprocal exchange. Evolution and Human Behavior, 28(3), 193-198.

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

Strimling, P., Enquist, M., Eriksson, K. (2009). Repeated learning makes cultural learning unique. PNAS, 106(33), 13870-13874.

Strohmetz, D. B., Rind, B., Fisher, R., and 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., and Flament, C. (1971). Social categorization and intergroup behaviour. European Journal of Social Psychology, 1, 149-178.

Takahashi, N., and 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.

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

Tooby, J., and 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., and 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.

Van den Bergh, B., and Dewitte, S. (2006). The robustness of the "Raise-the-Stakes" strategy: Coping with exploitation in noisy Prisoner's Dilemma Games. *Evolution and Human Behavior*, 27(1), 19-28.

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.

Vyrastekova, J., and van Soest, D. (2008). On the (in)effectiveness of rewards in sustaining cooperation. *Experimental Economics*, 11(1), 53-65.

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

Wedekind, C., and Braithwaite, V. A. (2002). The long-term benefits of human generosity in indirect reciprocity. *Current Biology*, 12, 1012-1015.

Wedekind, C., and Milinski, M. (2000) Cooperation through image scoring in humans. *Science* 288, 850-852.

West, S.A., Gardner, A., Shuker, D.M., Reynolds, T., Burton-Chellow, M., Sykes, E.M., Guinnee, M.A., and Griffin, A.S. (2006). Cooperation and the scale of competition in humans. *Current Biology*, 16, 1103-1106.

West, S. A., Griffin, A. S., and Gardner, A. (2007). Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, 20, 415-432.

Wheatcroft, D. J., and Price, T. D. (2008). Reciprocal cooperation in avian mobbing: playing nice pays. *Trends in Ecology and Evolution*, 23(8), 416-419.

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

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

Willer, R. (2009). Groups reward individual sacrifice: the status solution to the collective action problem. *American Sociological Review*, 74, 23-43.

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., and 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. (2005). Comparisons of Australians and Japanese on group-based cooperation. *Asian Journal of Social Psychology*, 8, 173-190.

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

Yang, C.-L., Yue, C.-S. J., and Yu, I.-T. (2007). The rise of cooperation in correlated matching prisoners dilemma: an experiment. *Experimental Economics*, 10(3), 3-20.

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 and C. Perrins), pp. 253-259. Baltimore, MD: University Park Press.

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

# INDEX